

# nature



THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

*Closing in on the  
long-sought  
metallic atomic  
state of hydrogen*

PAGE 63

## PRESSURE POINT

### PHYSICS

#### THAT'S LIFE

*Birds, cells and a  
theory of everything*

PAGE 16

### ALTERNATIVE ENERGY

#### MAKE WIND WORK

*Industry must share data  
to increase efficiency*

PAGE 19

### CULTURE

#### THE HOT TICKETS 2016

*Must-see exhibitions,  
music, plays and movies*

PAGE 22

NATURE.COM/NATURE

7 January 2016 £10

Vol. 529, No. 7584



# THIS WEEK

## EDITORIALS

**WORLD VIEW** Take a lesson in climate diplomacy from the Pope **p.6**

**CLEAN START** US regulators target Volkswagen over emissions **p.7**



**DESCRIPTION** Nobel-winning pharmacologist Alfred Gilman dies **p.7**

## Come together

*Cross-continent collaboration in the sciences has become the norm. We must ensure that disadvantaged regions are not left out.*

Amid the pledges to exercise and to keep a tidier office or bench space, scientists who wish to get on in 2016 should make a simple resolution for the new year: broaden your horizons. Think beyond the conventional format of the academic paper and experiment with new ways to present data and results. Look past the historical boundaries between academic subjects to the emerging landscape of interdisciplinarity. And, perhaps most importantly, embrace the growing trend of international collaboration.

The benefits of international partnership are clear. Cross-border research receives more attention than does insular work and its publications attract more citations. The promise to global science is obvious, too: publicly funded research increasingly looks for impact and pay back, and many of the most immediate problems that science can help with are not defined by national borders.

Issues of sustainability, health, access to food and water, stable ecosystems — the ‘grand challenges’ — are the products of complex chains and relationships, natural causes and human effects, across diverse yet connected regions. Solutions, and the science to seek these solutions, must sprout from a similar network: diverse yet connected.

The *Nature Index 2015 Collaborations* supplement published in November demonstrates the trend towards collaboration (see [go.nature.com/nji2gb](http://go.nature.com/nji2gb)). Some 70% of the academic papers analysed from the University of Cambridge, UK, for example, featured a co-author from a different country. It also demonstrates the shifting foundations for these international projects, which no longer need to be anchored to the usual big players of Europe’s leading lights, Japan and the United States. Scientists in Spain and Portugal are forging productive alliances with colleagues in South America. Australian researchers are increasingly looking to team up with scientists in the Asia-Pacific region.

This reflects the new, broader geopolitics of the twenty-first century — a change neatly illustrated by the climate-change agreement signed in Paris last month. Nations such as China, India and Brazil — previously defined in climate talks as poor developing countries — have taken on a more equal share of the responsibility for the struggle against global warming, to match their emerging higher status.

### ON THE OUTSIDE LOOKING IN

Not all scientists are benefiting from this era of cooperation. And, as bibliometrics specialists Jonathan Adams and Tamar Loach wrote in the *Nature Index* supplement, the cost of missing out can be severe (see J. Adams and T. Loach *Nature* **527**, S58–S59; 2015). “If collaboration is linked with high impact, then research groups who are not part of the collaborative network risk being left behind, marginalized by a lack of access to the cutting edge of research in their field.”

Where these excluded scientists live and work will come as no surprise. Africa remains under-represented in this new world, more heavily so if the relatively strong part played by South Africa compared with the rest of the continent is taken into account. Yet challenges

do not come much larger than those experienced in the patchwork of political, social and economic systems that make up the African continent. And as the Ebola virus outbreak has demonstrated, the problems of Africa — as well as having immediate and devastating local impacts — also challenge the rest of the world.

How can research and the growing strength of international collaboration reach more developing nations? How can we ensure that the products of scientific research reach the bulk of humanity who would benefit the most?

**“The long-term solution to inequality of opportunity is equality of investment.”**

It is no coincidence that China’s arrival on the global scene and as a desired partner comes on the tail of massive domestic investment in research. Many nations in Africa (and elsewhere in what is known as the global south) cannot or do not want to put serious money into science, and academic market forces — like it or not — will continue to drive parties in the global north elsewhere in search of synergies.

Instead, scientific investment by rich nations in poor countries and regions has long been tied to the development agenda. As such, it is, rightly, not judged on scientific output — papers and citations — alone. But alliances of unequal partners can be notoriously awkward, and so it has proved with research funded in this way. Post-colonial paternalism gave way to scientific aid, but that change did not challenge the donor-recipient dynamic and the polarizing problems it sets up in projects and relationships. In this model, those from the north who pay the bills too often decided the research agenda and how success will be defined, and those from the south were too often expected to fit in, provide the data and be grateful for the opportunity.

Plenty of players — from government funders and philanthropic bodies to institutions and individual project leaders — are taking admirable steps to call attention to this kind of inequality and to address it. Those efforts deserve praise and support.

The long-term solution to inequality of opportunity is equality of investment. For now, researchers involved with such asymmetric collaborations must ensure that they do not take advantage. As horizons expand, so must the professional codes and ethical safeguards that reward input with appropriate credit and govern the fair and equitable use of data and materials.

There must also be broader awareness that, just as there is more to research than papers, there is more provided to a partnership than conventional resources such as cash and equipment. The *Nature Index* supplement profiled an international project that published a genetic analysis of humans, chimpanzees and their lice. It quoted a Ugandan author on the paper as saying that it would have been impossible without the support of research partners in the United States and Europe, because the Ugandan group did not have the necessary technology. That is true. But then the partners did not have the necessary chimps. ■





## Constructive engagement is the key to climate action

*This year, scientists should resolve to follow the lead of Pope Francis and seek an inclusive approach to climate change, says Daniel Sarewitz.*

In 1992, the Catholic Church formally acknowledged that Galileo was right — Earth really does move around the Sun. That step towards reconciling religious dogma with science took around 380 years. What will 2016 bring? Whisper it, but science and the church seem to be walking hand-in-hand on one of the defining issues of the twenty-first century, and in a way that is truly remarkable.

The issue is climate change, and the force behind this new reconciliation is Pope Francis. He termed last month's COP21 global warming conference in Paris a "now or never" opportunity, and greeted the news that the talks had led to an international agreement by exhorting "the whole international community to proceed on the path undertaken in the name of an ever more effective solidarity".

Such sentiments built directly on the Pope's already famous encyclical letter of May 2015: a long and somewhat rambling critique of modernity that, among other things, called for changes of lifestyle, production and consumption to combat climate change.

Leading scientists have welcomed the Pope and the church into the fold of rationality. Johan Rockström, for example, the lead author of a widely cited article in *Nature* on planetary boundaries (J. Rockström *et al.* *Nature* **461**, 472–475; 2009), has noted approvingly: "Pope Francis' encyclical suggests — in line with our analysis — that planetary stewardship must now be the foundation of our values, beliefs and economic systems."

But the church's concern about climate is remarkable not so much in how it lines up with scientists' views, but in how it potentially challenges them. Indeed, the Pope's moral logic makes clear that widespread agreement on the science is much less important than is a political environment that welcomes diverse belief systems.

The effectiveness of the COP21 agreement — heralded as an important breakthrough because all nations have signed up to its aims — will depend on the ability of individual nations to reduce carbon emissions while still advancing the well-being of their citizens. This means that progress on climate will hinge on political decisions about how best to pursue both goals. Effective politics will, in turn, demand constructive engagement among multiple voices to achieve solutions that all can live with. "Solutions," writes the Pope, "will not emerge from just one way of interpreting and transforming reality."

Yet the original sin of climate-change policy in the United States was that from the beginning it ruled out such pluralism, because scientists and environmental activists alike tended to frame action in a way that could only alienate economic and social conservatives. Political rhetoric and policy proposals focused on demands for a global governance regime, government manipulation of markets and regulatory incentives for massive behaviour

change. From the perspective of US conservatives, it would be hard to imagine a more toxic combination of policy ambitions. And because scientists and climate activists claimed that science dictated their policy agenda, conservatives had every reason to be suspicious about the motives of the scientists and the credibility of their science. The legacy of that strategy is evident in the uniform scepticism of the Republican presidential candidates about global warming.

The Pope, however, draws direct connections between action on climate and conservative US touchstones such as 'family values'. He emphasizes the family as "the basic cell of society" and the starting place for action, because that is where "we first learn how to show love and respect for life" as well as "respect for the local ecosystem and care for all creatures". And because, as ecologically minded people often observe, every-

thing is connected to everything else, the Pope reasons that "concern for the protection of nature is also incompatible with the justification of abortion. How can we genuinely teach the importance of concern for other vulnerable beings, however troublesome or inconvenient they may be, if we fail to protect a human embryo?"

These values are likely to make many scientists and climate activists squirm. But, for the optimism stirred by the COP21 agreement to translate into tangible progress, climate politics in the United States will have to offer a serious place in the debate for the fundamental values that lie behind conservatism. As the Pope has shown, such values are perfectly compatible with action on climate. Along the way, they may even help to expose some of the contradictions and incoherence of

the mainstream climate-change regime.

For example, climate politics in the United States has often played out as a de facto attack on the cultural iconography of conservative middle America, such as pick-up trucks and muscle cars. Meanwhile, as people concerned about the climate jet off to international conferences and ecotourism sites, they can mitigate their guilt with carbon offsets — a modern sort of indulgence that the Pope terms a "ploy which permits maintaining the excessive consumption of some countries and sectors".

In acknowledging the climate problem, the Pope has also shown that in this new year, conservative voices and belief systems can begin to enter constructively into the climate debate after an absence of two decades. Yet, if science and religion are beginning to walk together, the devil remains in the politics. And this is where, logically enough, science can learn a thing or two from religion. ■

**Daniel Sarewitz** is co-director of the Consortium for Science, Policy and Outcomes at Arizona State University, and is based in Washington DC.  
e-mail: [daniel.sarewitz@asu.edu](mailto:daniel.sarewitz@asu.edu)

**IF SCIENCE AND  
RELIGION  
ARE BEGINNING  
TO WALK TOGETHER,  
THE DEVIL  
REMAINS IN THE  
POLITICS.**

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/r5zjy2](http://go.nature.com/r5zjy2)

# SEVEN DAYS

The news in brief

TIM SHARP/AP

## EVENTS

### Periodic addition

Four new elements have been officially added to the periodic table, completing its seventh row. The International Union of Pure and Applied Chemistry in Research Triangle Park, North Carolina, announced on 30 December that evidence supporting the discoveries of elements 113, 115, 117 and 118 by laboratories in Russia, the United States and Japan was valid. See [go.nature.com/vgug27](http://go.nature.com/vgug27) for more.

### Volkswagen sued

The US Environmental Protection Agency (EPA) is taking Volkswagen to court after revelations that the company fitted vehicles with devices that circumvent emissions regulations. Volkswagen has admitted using such 'defeat devices' and has apologized for fitting them to some models of Volkswagen, Audi and Porsche cars. On 4 January, the Department of Justice, acting on behalf of the EPA, said that nearly 600,000 vehicles sold in the United States had used

the illegal devices, causing harmful air pollution and violating the US Clean Air Act.

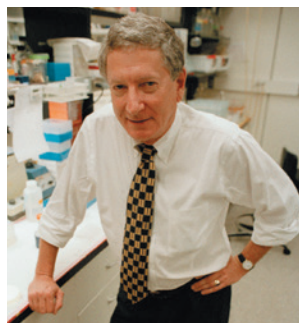
### Guinea Ebola-free

Ebola virus is no longer spreading in Guinea, the World Health Organization (WHO) declared on 29 December. The announcement came 42 days after the West African country's last patient, a newborn, tested negative for the virus for a second time. Health officials will now watch closely for flare-ups of the deadly disease. Last November, a cluster of three Ebola cases emerged in Liberia, months after the WHO had announced the end of Ebola transmission there.

## PEOPLE

### Alfred Gilman dies

Pharmacologist Alfred Goodman Gilman (pictured), who shared the 1994 Nobel Prize in Physiology or Medicine for his discovery of G proteins, died on 23 December, aged 74. G proteins are attached to the internal surface of a cell's membrane and are involved



in transmitting signals from the outside to the cell's interior. Around 40% of pharmaceuticals act by binding to specific receptors that are coupled to G proteins. Gilman was an editor of the textbook *The Pharmacological Basis of Therapeutics*, originally co-written by his father.

### Fellowship refusal

The American Association for the Advancement of Science said on 22 December that it will not award an honorary fellowship to chemist Patrick Harran, who was prosecuted for the accidental death of 23-year-old researcher Sheharbano Sangji in his lab in 2009. Last November, Harran was named as one of 347 scientists elected to

receive the honour. But the organization subsequently learnt of the death, from a chemical fire in Harran's lab at the University of California, Los Angeles, and reconsidered his nomination. See [go.nature.com/jsujun](http://go.nature.com/jsujun) for more.

## FACILITIES

### Moon base no more

Russia's plans to build a Moon base are on hold, according to 29 December reports. The Russian newspaper *Izvestia* says that a draft revised programme to 2025, developed by the country's space agency Roscosmos, no longer includes plans to create a lunar base, a long-held goal. The agency proposes cutting the budget for human Moon missions by 20%, or 88.5 billion roubles (US\$1.2 billion), the newspaper adds. Roscosmos told the news agency Reuters that it was revising the scale of its programme, but declined to comment on the figures.

## BUSINESS

### India biotech boost

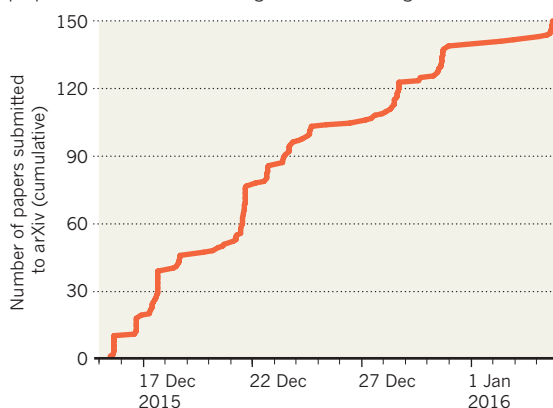
India has launched a strategy for biotechnology development for 2015–20, aiming to increase its biotech turnover from US\$7 billion to \$100 billion by 2025. The country will invest in a new generation of biotech products, create infrastructure for research and development and commercialization, and establish India as a major biomanufacturing hub, science and technology minister Harsh Vardhan announced on 30 December. Plans include a network of biotech incubators, technology-development centres and 150 technology-transfer organizations.

## TREND WATCH

Theoretical physicists are rapidly churning out papers as they rush to analyse tantalizing hints of a new particle — a boson — in data from the Large Hadron Collider (LHC). Experimental results announced on 15 December at CERN, which hosts the LHC near Geneva, Switzerland, sparked a flood of papers posted on the preprint server arXiv — 150 had been published as *Nature* went to press — even though the statistical significance of the findings is low. See [go.nature.com/eqmchr](http://go.nature.com/eqmchr) for more.

### HINT OF NEW BOSON SPARKS FLOOD OF PAPERS

In just 21 days, physicists have posted 150 papers on the arXiv preprint server about tantalizing results at the Large Hadron Collider.



SOURCE: PAUL GINSBURG/ARXIV

► NATURE.COM

For daily news updates see:

[www.nature.com/news](http://www.nature.com/news)



# NEWS IN FOCUS

**DISEASE** Guinea-worm eradication slowed by epidemic in dogs **p.10**

**CLIMATE** Antarctic clouds studied for first time in decades **p.12**

**PREVIEW** The science to look out for in the coming twelve months **p.14**

**PHYSICS** Flocking birds, swarming molecules and the mathematics of life **p.16**



FERNANDO MOLERES/PANOS PICTURES

people in different ways. By enabling physicians to target drugs only to those who will benefit, such knowledge can cut waste, improve health outcomes using existing treatments, and inform drug development. For example, it is now clear that individuals with a certain mutation (which is mostly found in Asian people) respond better to the lung-cancer drug Tarceva (erlotinib; W. Pao *et al. Proc. Natl Acad. Sci. USA* **101**, 13306–13311; 2004), and the discovery of a mutation that causes 4% of US cystic fibrosis cases led to the development of the drug Kalydeco (ivacaftor).

The Chinese government is expected to officially announce the initiative after it approves its next five-year plan in March. Just how much the effort will cost is unclear — but it will almost certainly be larger and more expensive than the US\$215-million US initiative.

Since last spring, Chinese media has been abuzz with estimates of a 60-billion yuan (US\$9.2-billion) budget, spread over 15 years. But this figure is not finalized, cautions Zhan Qimin, director of the State Key Laboratory of Molecular Oncology at Peking Union Medical College in Beijing, who is involved in the initiative. He says that the effort will consist of hundreds of separate projects to sequence genomes and gather clinical data, with support for each ranging from tens of millions of yuan to more than 100 million yuan.

Anticipating the initiative, leading institutes — including Tsinghua University, Fudan University and the Chinese Academy of Medical Sciences — are scrambling to set up precision-medicine centres. Sichuan University's West China Hospital, for instance, plans to sequence 1 million human genomes itself — the same goal as the entire US initiative. The hospital will focus on ten diseases, starting with lung cancer.

Both the US and the Chinese efforts will focus on genetic links to diseases that are particularly deadly, such as cancer and heart disease. But China will target specific cancers, such as stomach and liver cancer, which are common there.

The Chinese initiative is part of a series of research-funding efforts that will replace two major grant programmes, known as 863 and 973, that are due to be phased out by 2017. The new programmes will be “more organized, more efficient”, says Zhan.

Genome-sequencing companies are already vying to provide services to deal with the anticipated demand. For several years, China has boasted high genome-sequencing capacity. ►

Precision medicine uses genomic and physiological data to tailor treatments to individuals.

## HEALTH CARE

# China embraces precision therapy

*Strong genomics record bodes well for health-care revolution.*

BY DAVID CYRANOSKI

**F**ormidable capacity in genome sequencing, access to millions of patients and the promise of solid governmental support: those are the assets that China hopes to bring to the nascent field of precision medicine, which uses genomic, physiological and other data to tailor treatments to individuals.

Almost exactly one year after US President Barack Obama announced the Precision

Medicine Initiative, China is finalizing plans for its own, much larger project. But as universities and sequencing companies line up to gather and analyse the data, some observers worry that problems with the nation's health-care infrastructure — in particular a dearth of doctors — threaten the effort's ultimate goal of improving patient care.

Precision medicine harnesses huge amounts of clinical data, from genome sequences to health records, to determine how drugs affect

► In 2010, the genomics institute BGI in Shenzhen was estimated to host more sequencing capacity than the entire United States. This was thanks to its equipment, purchased from Illumina of San Diego, California, which at the time represented state-of-the-art technology. But Illumina has since sold upgraded machines to at least three other genomics firms — WuXi PharmaTech and Cloud Health, both in Shanghai, and the Beijing-based firm Novogene.

Jason Gang Jin, co-founder and chief executive of Cloud Health, says that this trio, rather than BGI, will be the main sequencing support for China's precision-medicine initiative — although BGI's director of research, Xu Xun, disagrees. Xu says that precision medicine is a priority for BGI and that the organization has a diverse portfolio of sequencers that still gives it an edge. "If you are talking about real data output, BGI is still leading in China, maybe even globally," he says. BGI has already established a collaboration with the Zhongshan Hospital's Center for Clinical Precision Medicine in Shanghai, which opened in May 2015 with a budget of 100 million yuan and is run by Fudan University.

#### NUMBERS GAME

Regardless of the details, Jin thinks that China will be faster than the United States at sequencing genomes and identifying mutations that are relevant to personalized medicine because China's larger populations of patients for each disease will make it easier to find sufficient numbers to study.

Still, it remains to be seen whether China has the resources to apply these insights to the individualized care of patients. "China wants to do it, and everybody is very excited," says Ta Jen Liu, project director at the MD Anderson Cancer Center in Houston, Texas, who helps to establish collaborations in China and is familiar with the precision-medicine scene there.

But there are hurdles. He notes that Chinese researchers and pharmaceutical companies have not had much success in developing drugs so far; that the pathologists needed to diagnose specific diseases are scarce in China; and that physicians there are notoriously overworked. "Doctors are always overwhelmed with patients, seeing 60 or 70 a day," he says. "They don't have time to sit down and think about what is best for specific patients."

David Weitz, a physicist at Harvard University who is starting a company in Beijing to develop diagnostic instruments for use in precision medicine, agrees that there will be obstacles, but notes the initiative's assets. "We need lots of data to validate ideas, to validate tests," he says. "There's lots of data here."

He thinks that this, combined with the Chinese government's determination to succeed, will mean that the effort will ultimately win out. "They really seem devoted to meeting the needs of the society," he says. "It's an exciting thing, to try to help that many people." ■



Most cases of Guinea-worm disease in Chad have occurred in communities based along the Chari River.

#### INFECTIOUS DISEASE

# Dogs thwart end to Guinea worm

*Epidemic in dogs complicates push to wipe out parasite.*

BY EWEN CALLAWAY

A decades-long push to make Guinea-worm disease the first parasitic infection to be wiped out is close to victory. But a mysterious epidemic of the parasite in dogs threatens to foil the eradication effort.

"If we're going to be aggressive and achieve this, we have to eliminate the infection in dogs," says David Molyneaux, a parasitologist at the Liverpool School of Tropical Medicine, UK.

The Carter Center in Atlanta, Georgia, is leading the global campaign to eradicate Guinea worm. Next week, it will announce that case numbers for the excruciatingly painful infection are at a record low, with approximately 25 cases reported in 2015 in just 4 countries: Chad, Ethiopia, Mali and South Sudan. But infections in dogs are soaring in Chad, where officials will meet at the end of January to grapple with the canine epidemic. The central African nation recorded more than 450 cases of Guinea worm in domestic dogs last year — an all-time high (see 'Canine comeback').

Researchers and officials strongly suspect that dogs are spreading the infection to humans; now the race is on to understand how this might happen, as well as how dogs

acquire the infection in the first place. The World Health Organization is unlikely to declare Guinea worm eradicated until the parasite has stopped spreading in dogs, says Molyneaux, who is part of the commission that will make that decision.

In 1986, when the Carter Centre joined the Guinea-worm eradication campaign, there were an estimated 3.5 million infections annually, mostly due to poor sanitation and lack of access to clean water.

When people drink unfiltered water, they can swallow microscopic freshwater crustaceans called copepods, which Guinea-worm larvae infect. The copepods die, releasing the larvae, which mature and mate in the human intestine. Male worms die after mating, but adult females — approximately 80 centimetres in length — survive and slowly migrate out of the gut. About a year after infection, they burrow through their host's skin, usually around the legs and feet, sometimes taking weeks to fully escape. To cope with the searing pain, many people bathe in rivers and lakes, contaminating the water with the next generation of larvae. Although rarely fatal, Guinea worm can debilitate people for months and keep children out of school.

There is no vaccine against the parasite and no effective treatment, so eradication





PHILIPPE DESMAZES/AFP/GETTY

efforts have focused on providing clean water and changing people's behaviour, says Donald Hopkins, a special adviser at the Carter Center who is leading its Guinea-worm eradication efforts. People in areas in which the parasite was once rife have learnt to filter their water using cloths and to avoid re-contaminating water supplies. Even the most out-of-the-way villages now quickly contain cases and report them to health officials.

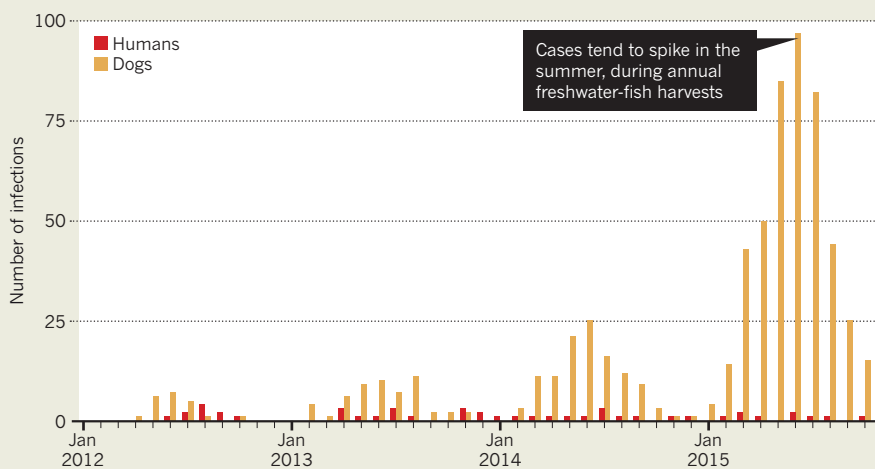
Chad was on the cusp of being declared free of Guinea worm in the late 2000s: no case had been recorded in the previous decade. But starting in April 2010, increased surveillance turned up a handful of human infections, and around 60 cases have been recorded since then.

The cases are unusually sporadic and isolated from one another, says Mark Eberhard, a parasitologist who consults on Guinea-worm eradication for the Carter Center. More typically, cases occur in clusters and recur in the same village year after year. "There was no increase or explosion of cases as one would expect," he says.

Shortly after these observations, officials began to hear rumours of Guinea-worm-infected dogs in Chad. Researchers have known for decades that dogs, leopards and other mammals occasionally acquire Guinea-worm-like infections, but they assumed that these cases stemmed from distinct species of *Dracunculus*, the nematode worm that causes the disease, or were rare examples of infections

## CANINE COMEBACK

The number of Guinea-worm cases in dogs is soaring in Chad — a development that threatens global efforts to eradicate the parasite.



SOURCE: CARTER CENTER/CDC

that had somehow spilt over from an outbreak in humans.

But in Chad, researchers now think that dogs are spreading the worms to humans — not the other way around. Between January and October 2015, officials recorded 459 canine infections from 150 villages in the central African nation — an unprecedented volume. And genome sequencing has confirmed that dogs in Chad are infected by the same nematode worms (*Dracunculus medinensis*) that plague humans (M. L. Eberhard *et al.* *Am. J. Trop. Med. Hyg.* **90**, 61–70; 2014).

To better understand the situation, a team led by James Cotton and Caroline Durrant, genome scientists at the Wellcome Trust Sanger Institute in Hinxton, UK, is now sequencing the genomes of more Guinea worms collected from dogs and humans in Chad to confirm that dogs are indeed transmitting the disease to people. And Eberhard, who is convinced that this is the case, is trying to determine how dogs become infected in the first place. They are unlikely to contract the worms from drinking water, he says, because dogs tend to scare away copepods when they lap. Most of Chad's cases have occurred among fishing communities along the Chari River, and Eberhard

suspects that dogs are eating the entrails of gutted, copepod-eating fish. Dogs then pass the worms to humans by reintroducing the larvae into water.

Researchers, including Eberhard, are testing aspects of this hypothesis in ferrets, a common animal model in disease research, but eradication officials in Chad are not waiting for the results before taking action. Since February 2015, they have offered the equivalent of US\$20 to people who report Guinea-worm cases in dogs and tie up the animals to prevent them from contaminating water sources. They are also encouraging villagers to bury fish entrails to keep dogs from eating them. And a trial is ongoing to test whether a drug used to treat heartworm — a roundworm parasite common in dogs — has any effect on Guinea worm. Because of Guinea worm's one-year incubation time, it should be clear before the end of 2016 whether these interventions have worked.

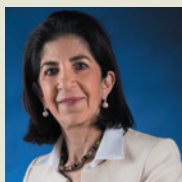
Older residents from villages along the Chari River say that their fishing practices have not changed, according to Hopkins, and they cannot recall dogs becoming infected with Guinea worm in the past. But Molyneux says that the dearth of humans transmitting the disease could explain the parasite's jump to dogs. "If you were Guinea worm and there were only 100 of you left in the world," he says, "what would you do? You'd get the hell out of the host that's being targeted and move to something else." ■

MAXIMILIEN BRICE/CERN



## MORE ONLINE

### MORE ONLINE



Q&A with CERN's new director-general, Fabiola Gianotti  
[go.nature.com/pxdcxd](http://go.nature.com/pxdcxd)

### MORE NEWS

- Four chemical elements added to periodic table [go.nature.com/frr5xt](http://go.nature.com/frr5xt)
- Scientific society withholds honour over lab death [go.nature.com/jsujun](http://go.nature.com/jsujun)
- Missing mice, data and protocols revealed in biomedical studies [go.nature.com/ip9flo](http://go.nature.com/ip9flo)

### NATURE PODCAST



Science in 2016; how extreme weather affects crops; and a new phase of hydrogen  
[nature.com/nature/podcast](http://nature.com/nature/podcast)



Clouds in Antarctica can influence weather the world over.

CLIMATE SCIENCE

# Antarctic cloud study takes off

*Scientists probe atmospheric physics above ice sheet for the first time since the 1960s.*

BY ALEXANDRA WITZE

On Antarctica's Ross Island, a short drive from the US McMurdo research station, high-tech radar antennas and other atmospheric instruments gaze skyward, gathering detailed measurements of West Antarctic clouds. Remarkably, these are the first such data to be gathered in five decades — even though weather patterns in the region can influence those half a world away.

The US\$5-million project, known as the Atmospheric Radiation Measurement West Antarctic Radiation Experiment (AWARE), began to observe the skies near McMurdo in November and will run until early 2017. A second measurement station, 1,600 kilometres away in the ice sheet's interior, will operate until the end of this month. (The site is so remote that it can be used only during the Antarctic summer.)

A similar experiment in the Arctic in 1997–98 relied on an instrument-laden ship that was deliberately frozen into sea ice. It yielded fundamental insights into the physics of northern polar clouds<sup>1</sup>, and AWARE scientists hope that their project will do the same

for the south. “This is going to be a sea change in our understanding,” says Lynn Russell, an atmospheric scientist at the Scripps Institution of Oceanography in La Jolla, California, and a co-principal investigator on AWARE.

Antarctica's massive ice sheet acts as a global heat sink. As a result, changes in Antarctic clouds, such as the amount of ground they cover or how much radiation they absorb, can have ripple effects as far away as the tropics. Climate modellers need to understand the physics of these clouds if they are to correctly work out how weather around the globe will change as the polar regions warm.

Scientists have not made detailed, in-place measurements of the skies above West Antarctica since 1967, when weather-balloon launches ceased a decade after they began during the 1957–58 International Geophysical Year, says Russell.

AWARE, which is led by Scripps atmospheric scientist Dan Lubin, aims to get the best data yet on clouds and aerosol particles above

West Antarctica. That includes mixed-phase clouds, which occur in polar regions and combine supercooled water with ice. Studies have shown that clouds moving across Antarctica's interior are mostly ice, whereas those moving onshore from the coast contain more liquid water<sup>2</sup>. The composition of these clouds plays a major part in determining how much sunlight they reflect into space — which helps to shape atmospheric circulation and weather patterns below.

Satellites such as NASA's CloudSat and CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations) can probe the internal structure of Antarctic clouds<sup>3</sup>, but in only a narrow ribbon as seen directly beneath the spacecraft's orbit. AWARE uses multiple radar instruments and a sophisticated lidar system to explore the clouds' many layers, examining properties such as phase and particle size at various altitudes.

Early AWARE data show mixed-phase clouds over McMurdo, in the first detailed measurements of such cloud systems outside the Arctic. “The Antarctic is a very different environment than the Arctic, because it is colder year-round and also has a very pristine atmosphere,” says Lubin. Team scientists reported early results on 16 December at the American Geophysical Union meeting in San Francisco, California.

The team has also clocked pulses of humidity swinging in and out of the McMurdo area as a storm passed through, altering how the clouds transmit radiation.

Getting the basic data should help scientists to better understand how Antarctic clouds will respond to a changing climate, Russell says. West Antarctica is warming by as much as 0.4°C per decade, and as its ice melts, sea levels will rise. The AWARE measurements from the West Antarctic interior are designed to capture the height of the summer melt season.

One major question is how climate change may be intensifying westerly winds around Antarctica, and what those changes will do to southern polar clouds, says Andrew Vogelmann, an atmospheric scientist at Brookhaven National Laboratory in New York. With one AWARE location near the coast and another in the interior, project scientists aim to compare how atmospheric systems passing through West Antarctica affect both locations, and how those changes translate to wider global shifts.

One final twist, Vogelmann adds, is the presence this year of the El Niño weather pattern, which could affect conditions at the poles. “We may be able to catch some of that,” he says. ■

1. Intrieri, J. M., Shupe, M. D., Uttal, T. & McCarty, B. J. *J. Geophys. Res.* <http://dx.doi.org/10.1029/2000JC000423> (2002).
2. Scott, R. C. & Lubin, D. *J. Geophys. Res. Atmos.* **119**, 6702–6723 (2014).
3. Adhikari, L., Wang, Z. & Deng, M. *J. Geophys. Res.* **117**, D04202 (2012).

JASON EDWARDS/NATL GEOGRAPHIC CREATIVE



## PUBLISHING

# Dutch lead European push to flip journals to open access

*Academic consortia urge faster changes in scholarly publishing.*

BY DECLAN BUTLER

The Netherlands is leading what it hopes will be a pan-European effort in 2016 to push scholarly publishers towards open-access (OA) business models: making more papers free for all users as soon as they are published.

In 2014, publishers worldwide made 17% of new papers OA immediately on publication, up from 12% in 2011 (see 'Growth of open access'). But most papers are still locked behind paywalls when they are first published. The Dutch government, which took over the six-month rotating presidency of the European Union council of ministers this month, has declared furthering OA to be one of its top priorities.

With strong support from Carlos Moedas, the EU's research commissioner, it is planning a series of discussions on the issue — between European science ministers at the end of January (with a keynote talk from Bill Gates, whose philanthropic foundation strongly supports OA) and at an EU presidency conference on open science in April. At that forum, the European Commission is expected to launch an 'Open Science Policy Platform' with a remit that includes investigating how subscription publishers can best transition to OA.

The Association of Universities in the Netherlands (VSNU), a consortium of 14 institutes, has already taken radical steps. With backing from the Dutch government, it has negotiated several deals with major publishers over the past two years to make more Dutch papers open in subscription journals, with the aim of shifting the journals to an OA business model. The deals are a "great step forward to an OA world", says Paul Ayris, head of library services at University College London and a spokesperson for the League of European Research Universities, which has urged the commission and the Dutch presidency to speed the OA transition.

## OPEN-ACCESS DEALS

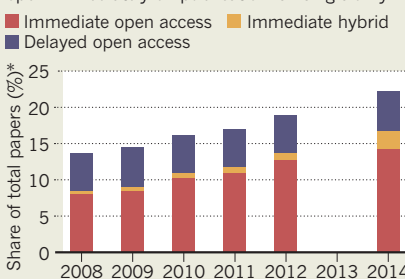
In 2014, the VSNU announced a deal in which it renewed its subscription to a bundle of 2,000 paywalled journals from the publisher Springer, but with terms that made papers by corresponding authors at subscribing Dutch universities OA, for no extra charge. (Springer has since merged with *Nature's* publisher.) Shortly before Christmas 2015, the VSNU announced

a similar agreement with Elsevier, which the consortium had threatened to boycott if its demands were not met: by 2018, 30% of Dutch papers will be OA in VSNU-subscribed Elsevier journals.

The hope, Ayris says, is that if other nations' organizations can make similar deals, publishers will be compelled to release more open papers in return for their flow of subscription income, effectively flipping their journals to become fully OA. OA journals receive no subscription income and instead make money either by direct subsidy or by charging authors (or their research funders) a fee to publish each OA paper.

## GROWTH OF OPEN ACCESS

The worldwide share of papers that journals make open immediately on publication is rising slowly.



The United Kingdom has gone down the same track. In October 2015, Jisc, a non-profit body that represents UK higher-education institutions, negotiated a deal that made OA papers with UK-based corresponding authors free in 1,600 selected Springer subscription journals. A spokesperson for Springer says that the agreements are pilots, but that "deals which combine subscriptions with OA publishing could accelerate the transition to OA on a large scale".

## HYBRID CRACKDOWN

A major driving force for the Dutch and British deals was to combat the expensive and controversial 'hybrid' business models that have been adopted by many subscription journals worldwide. Hybrid journals collect subscriptions but allow authors to make individual papers open for a fee. They charge higher fees, on average, than do fully OA journals, yet scientists who want OA papers often choose to publish with them because they are

generally more established or prestigious than many recently launched OA journals.

Robert Kiley, who is head of digital services at the library of the Wellcome Trust, the London-based biomedical funder, notes that many UK organizations have each paid millions of pounds to hybrid journals for open papers — while also paying them subscriptions. A deal akin to the VSNU's one with Springer would help to bypass this hybrid market.

But these kinds of deals have their critics. The costs of the agreements are confidential, points out Mark McCabe, an economist at the University of Michigan in Ann Arbor; he surmises that they did not come cheap. He and others say that such secretive deals risk locking academic institutions into continuing to pay expensive fees to major subscription publishers, and they shield the latter from competition.

McCabe proposes a more radical strategy: libraries or university consortia should stop paying journal subscriptions and should transfer the money saved to their researchers, who can use it to publish OA in journals of their choice. That way, authors might become more sensitive to the price of publishing — which might lead to greater competition between journals, promoting leaner-run OA journals that charge lower fees.

Some funders are trying other ideas to support OA but steer researchers away from the hybrid market. The Norwegian Research Council and the German Research Foundation both pay OA fees for researchers but prohibit them from being spent on articles in hybrid journals. And the Austrian Science Fund has capped OA payments at a certain level; if researchers want to publish in more expensive journals (often the hybrids), they must find the extra cash themselves.

But measures to change industry business models will succeed only with international buy-in. And some other nations, such as the United States, have not followed the Netherlands in urging the publishing industry to make more papers immediately OA. They have favoured other routes to free-to-read papers, such as encouraging academics to archive their pre-publication manuscripts online, and mandating subscription publishers to make papers free after a delay (typically six months or a year after publication). A successful push for immediate OA, Kiley says, would ultimately need to be global — not limited to Europe. ■

## RESEARCH

# What to look out for in 2016

*Space missions, carbon capture and gravitational waves are set to shape the year.*

## SUCKING UP CO<sub>2</sub>

A Swiss company is set to become the first firm to capture carbon dioxide from the air and sell it on a commercial scale, a stepping stone to larger facilities that could one day help to combat global warming. Around July, Clime-works will start capturing some 75 tonnes of CO<sub>2</sub> per month at its plant near Zurich, then selling the gas to nearby greenhouses to boost crop growth. Another company — Carbon Engineering in Calgary, Canada, which has been capturing CO<sub>2</sub> since October but is yet to bring it to market — hopes to show that it can convert the gas into liquid fuel. Facilities worldwide already capture the gas from power-plant exhausts, but until 2015 only small demonstration projects sucked it up from air.

## CUT-AND-PASTE GENES

Human trials will get under way for treatments that use DNA-editing technologies. Sangamo Biosciences in Richmond, California, will test the use of enzymes called zinc-finger nucleases to correct a gene defect that causes haemophilia. Working with Biogen of Cambridge, Massachusetts, it will also start a trial to look at whether the technique can boost a functional form of haemoglobin in people with the blood disorder  $\beta$ -thalassaemia. Scientists and ethicists hope to agree on broad safety and ethical guidelines for gene editing in humans in late 2016. And this year could see the birth of the first gene-edited monkeys that show

symptoms of the human disorders they are designed to model.

## HIGH COSMIC HOPES

Physicists think there is a good chance that they will see the first evidence of gravitational waves — ripples in space-time caused by dense, moving objects such as spiralling neutron stars — thanks to the Advanced Laser Interferometer Gravitational-Wave Observatory (Advanced LIGO). And Japan will launch Astro-H, a next-generation X-ray satellite observatory that, among other things, could confirm or refute the claim that heavy neutrinos give off dark-matter signals known as bulbulons. Hints of a potential new particle from the supercharged Large Hadron Collider (LHC), which has been running at record energies since last June, could become clearer as the machine rapidly accumulates data. Even if the particle is not confirmed, the LHC could still unearth other exotic phenomena, such as glueballs: particles made entirely of the carriers of the strong nuclear force.

## RISKY RESEARCH

Scientists will soon hear whether funding for research that makes viruses more dangerous can resume. In October 2014, the US government abruptly suspended financial support for 'gain-of-function' studies. These experiments could increase understanding of how certain pathogens evolve and how they can be destroyed, but critics say that the work

also boosts the risk of, for example, accidental release of deadly viruses. A risk-benefit analysis was completed in December 2015, and the US National Science Advisory Board for Biosecurity will issue recommendations in the next few months on whether to resume funding — potentially with tightened restrictions on the research.

## TO MARS AND BEYOND

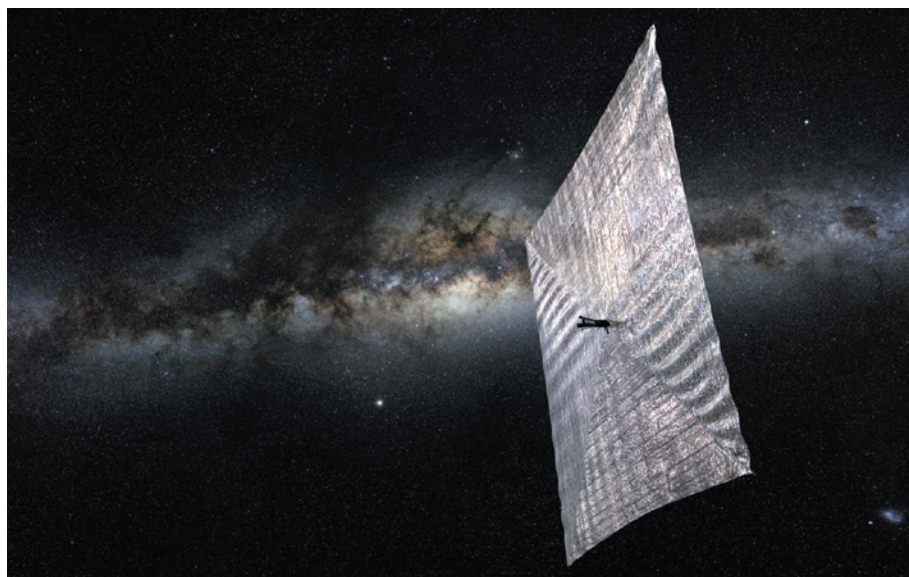
The orbits of Earth and Mars will bring the planets close to each other this year, creating the perfect opportunity for a trip to the red planet. A joint mission between the European Space Agency (ESA) and Roscosmos will capitalize on that chance. Launching in March, ExoMars 2016 will analyse gases in Mars's atmosphere and test landing technology. Farther afield, NASA's Juno mission will arrive at Jupiter in July. In September, ESA's craft Rosetta will make a death dive into the comet it orbits; mourners can console themselves with the launch of NASA's OSIRIS-REx, a mission to bring back samples from the asteroid Bennu.

## COMMERCIAL GAINS

One lucky research group will win a \$50-million grant for heart-disease research from Internet giant Google and the American Heart Association. Google's disease-research portfolio is growing, and neuroscientists are eager to see what Thomas Insel, former director of the US National Institute of Mental Health, will do at the firm, where he has been leading a mental-health effort since November. Private funding could also make its mark in space: the non-profit Planetary Society in Pasadena, California, plans to launch a US\$4.5-million mission in April to test its light-driven spacecraft, LightSail.

## SPACE DRIVE

Hot on the heels of the launch of the US\$100-million Dark Matter Particle Explorer (DAMPE) last December, China's National Space Science Center will launch the second and third space-science probes in its planned series of five. The world's first quantum communications test satellite will blast off in June, and the Hard X-ray Modulation Telescope — which will scour the sky for energetic sources of radiation, such as black holes and neutron stars — will fly by the end of the year. September will see China complete construction of the 500-meter Aperture Spherical Radio Telescope (FAST), which will supersede Puerto



The light-driven spacecraft LightSail will undergo a test mission in April.

JOSH SPRADLING/THE PLANETARY SOCIETY



Rico's Arecibo Observatory as the world's largest radio telescope. In Hawaii, the team behind the controversial Thirty Meter Telescope, which had its construction permit revoked in December, will try to work out whether and how it can move the project forward.

### MICROLIFE REVEALED

The first results from an ambitious project to analyse the world's microbial communities are expected this year. The Earth Microbiome Project, which launched in 2010, aims to sequence and characterize at least 200,000 samples of microbial DNA taken from everything from Komodo dragon tongues to soil in the Siberian tundra. The project promises to uncover unprecedented levels of biological diversity.

### POLITICAL UPHEAVAL

In November, the United States will elect a new president. If a Republican takes the White House, long-debated plans to bury nuclear waste at Yucca Mountain in Nevada may well resurface, and federal funding for climate and social science could face the chop.



Komodo dragon saliva has been sampled for the Earth Microbiome Project.

And if Canada's Liberal government lives up to its pre-election promises, the country will get a chief science officer, who researchers trust will arrive with a drive to rebuild the depleted ranks of government scientists.

### DREAM GENES

Neuroscientists hope to finally identify genes that are crucial to regulating the timing and duration of sleep but have been difficult to


tease out, possibly because they also have other functions in the brain. Pinpointing these genes could shed light on sleep disorders and some psychiatric illnesses, which scientists now realize are linked to highly disrupted sleep patterns.

### LET THERE BE LIGHT

The SESAME (Synchrotron-light for Experimental Science and Applications in the Middle East) facility will switch on in Jordan towards the end of 2016. The ring-shaped particle accelerator will generate intense light to probe materials and biological structures down to the atomic level. It is the region's first major international

research facility, and a rare collaboration between governments including Iran, Israel and the Palestinian Authority. Support to build a similar facility in Africa is likely to gather pace. And in June, scientists will get to use bright X-ray beams at the world's first fourth-generation synchrotron, MAX IV in Lund, Sweden. ■

COMPILED BY ELIZABETH GIBNEY



# The physics of life

*From flocking birds to swarming molecules, physicists are seeking to understand ‘active matter’ — and looking for a fundamental theory of the living world.*

BY GABRIEL POPKIN

**F**irst, Zvonimir Dogic and his students took microtubules — threadlike proteins that make up part of the cell’s internal ‘cytoskeleton’ — and mixed them with kinesins, motor proteins that travel along these threads like trains on a track. Then the researchers suspended droplets of this cocktail in oil and supplied it with the molecular fuel known as adenosine triphosphate (ATP).

To the team’s surprise and delight, the molecules organized themselves into large-scale patterns that swirled on each droplet’s surface. Bundles of microtubules linked by the proteins moved together “like a person crowd-surfing at a concert”, says Dogic, a physicist at Brandeis University in Waltham, Massachusetts.

With these experiments, published<sup>1</sup> in 2012, Dogic’s team created a new kind of liquid crystal. Unlike the molecules in standard liquid-crystal displays, which passively form patterns in response to electric fields, Dogic’s components were active. They propelled themselves, taking energy from their environment — in this case, from ATP. And they formed patterns spontaneously, thanks to the collective behaviour of thousands of units moving independently.

These are the hallmarks of systems that

physicists call active matter, which have become a major subject of research in the past few years. Examples abound in the natural world — among them the leaderless but coherent flocking of birds and the flowing, structure-forming cytoskeletons of cells. They are increasingly being made in the laboratory: investigators have synthesized active matter using both biological building blocks such as microtubules, and synthetic components including micrometre-scale, light-sensitive plastic ‘swimmers’ that form structures when someone turns on a lamp. Production of peer-reviewed papers with ‘active matter’ in the title or abstract has increased from less than 10 per year a decade ago to almost 70 last year, and several international workshops have been held on the topic in the past year.

## THE SECRET OF LIFE

Researchers hope that this work will lead them to a complete, quantitative theory of active matter. Such a theory would build on physicists’ century-old theory of statistical mechanics, which explains how the motion of atoms and molecules gives rise to everyday phenomena such as heat, temperature and pressure. But it could go much further,

providing a mathematical framework for still-mysterious biological processes such as how cells move things around, how they create and maintain their shapes and how they divide. “We want a theory of the mechanics and statistics of living matter with a status comparable to what’s already been done for collections of dead particles,” says Sriram Ramaswamy, a physicist and director of the Tata Institute of Fundamental Research’s Centre for Interdisciplinary Sciences in Hyderabad, India.

It could be a while before that want is satisfied, however. Experimentalists are only beginning to gain control of active materials in the lab. Even the most enthusiastic proponents of this research admit that no one has yet produced a theory of active matter that describes the behaviour of everything from cell parts to birds. And if such a theory did exist, it’s far from certain that mainstream biologists would see value in it. For biologists, the idea that living matter is active “would be just so obvious as to not really contain very much information”, says Jonathon Howard, a molecular biophysicist at Yale University in New Haven, Connecticut.

But that has not kept proponents from imagining applications such as self-assembling

HANS OVERDUIN/INIS/MINDEN/GETTY



## Flocking birds can synchronize to make patterns.

artificial tissue, self-pumping microfluidic devices and new bio-inspired materials — although researchers admit that such ideas are still far from being realized. “I think it’s too early for the field to have an application, because we’re still kind of astonished at what can happen,” says Andreas Bausch, a physicist at the Technical University of Munich in Germany — “but I do think the field needs somebody doing it.”

## ALL TOGETHER NOW

All known life forms are based on self-propelled entities uniting to create large-scale structures and movements. If this didn’t happen, organisms would be limited to using much slower, passive processes such as diffusion to move DNA and proteins around inside cells or tissues, and many of life’s complex structures and functions might never have evolved. Biologists and physicists have speculated for decades about the general principles of living matter, but research on cellular processes has focused on identifying the dizzying array of molecules involved, rather than on working out the principles by which they self-organize. As a result, what is now known as active-matter research did not really get under way until the mid-1990s.

One of the most influential early experiments was conducted by the team of Stanislas Leibler, a biophysicist who was then at Princeton University in New Jersey and is now at the Rockefeller University in New York. The group was among the first to show that complex, life-like structures could self-assemble from microtubules and a few proteins supplied with ATP<sup>2</sup>. Around the same time, an influential model of active matter was being developed by Tamás Vicsek, a theoretical biophysicist at Eötvös Loránd University in Budapest. In the early 1990s, Vicsek was trying to account for the collective motions of bird flocks, bacterial colonies and cytoskeleton components when he realized that no existing theory would work. “It’s not like equilibrium statistical mechanics, where you take a book and you find what to do,” says physicist Jean-François Joanny of the Curie Institute in Paris.

Instead, Vicsek found a starting point in a model of magnetic materials developed in 1928 by German physicist Werner Heisenberg. Heisenberg imagined each atom as a freely rotating bar magnet, and found that large-scale magnetism emerges when interactions between these atomic magnets cause the majority of them to align. To explain active matter, Vicsek replaced the tiny magnets with moving ‘arrows’ symbolizing particles with velocities that aligned with the average velocity of their neighbours — albeit with a certain amount of random error. That led to what is now known as Vicsek’s flocking model<sup>3</sup>. His simulations showed that when enough arrows were packed into a small enough space, they began to move in patterns that closely resembled the familiar movements of bird flocks and

fish schools (see ‘Smart swarm’).

“I got excited,” recalls Vicsek, whose 1995 paper<sup>3</sup> on the model has received more than 3,500 citations. “I was walking up and down the corridor and told people I had designed the moving version of the Heisenberg model.”

One physicist attracted to this idea was John Toner, who heard Vicsek give a talk on it in 1994. Toner, now at the University of Oregon in Eugene, saw that Vicsek’s swarming arrows could be modelled as a continuous fluid. He took the standard equations for hydrodynamics, which describe fluid flow in everything from tea kettles to oceans, and modified them to account for how individual particles use energy<sup>4</sup>. Toner’s fluid model and Vicsek’s discrete-particle model gave essentially the same predictions for a wide range of phenomena, and launched a cottage industry of active-matter simulations.

There was only one problem. Whereas the number of simulations was skyrocketing, says physicist Denis Bartolo of the École Normale Supérieure in Lyons, France, “the number of quantitative experiments was constant and very close to zero”. Practical work was challenging: no one could hope to do controlled experiments with 10,000 real birds or fish. And at the microscopic scale, few scientists were familiar with both the necessary theoretical work — being published mainly in physics journals — and the biological lab techniques needed to purify cellular components.

## PRACTICAL MAGIC

Only in the late 2000s did the theoretical and experimental pieces begin coming together. Bausch led one of the first precise, quantitative experiments. He and his colleagues mixed actin, a filament that forms most of the cytoskeleton of complex cells, with myosin, a molecular motor that ‘walks’ on actin and makes muscles contract. The researchers added myosin’s natural fuel, ATP, then put the mixture on a microscope slide and watched. “We didn’t do anything; we just added the stuff,” Bausch says. At low concentrations, the actin filaments swam around without recognizable order. But at higher densities, they formed pulsating clusters, swirls and bands. Bausch and his colleagues immediately recognized and quantified phase transitions of the kind that Vicsek and others had predicted. Their 2010 paper<sup>5</sup> helped to ignite the experimental active-matter field.

Among the studies that followed were Dogic’s 2012 microtubule experiments<sup>1</sup>, which used another walking protein, kinesin. The resulting patterns were much more complex and dynamic than the ones Bausch saw: the flowing microtubules looked like fingerprint whorls in motion. Dogic and his team also noticed that the orderly alignment of this flow would occasionally break down and produce ‘defects’: discontinuities in the pattern that resemble converging longitude lines at the North and South poles. These defects were dynamic, moving around like self-propelled particles.

No theory at the time could account for this behaviour. But in 2014, Dogic teamed up with Bausch and physicist Cristina Marchetti of Syracuse University in New York to describe the behaviour of active liquid crystals swirling on spherical vesicles in terms of the movement of defects rather than of individual crystal elements<sup>6</sup>. Furthermore, the group found that it could tune the defects’ motion by adjusting the vesicle’s diameter and surface tension, suggesting a possible way to control an active crystal.

Dogic and his students are now trying to do just that. By studying the spontaneous flows of microtubules and proteins confined in small, doughnut-shaped containers, they hope to lay the groundwork for a self-pumping fluid that could move molecules around in microfluidic devices similar to those that are becoming increasingly common in experimental biology, medicine and industry. Active matter “changes our ideas of what materials can do,” says Dogic.

But any industrial application will have to overcome at least one major roadblock. The biological materials currently used in active-matter experiments are expensive and time-consuming to purify — Dogic’s microtubules come from

cow brains, and Bausch uses actin from rabbit muscle — and they last only a short time in the lab. Until a cheap, robust, off-the-shelf source of active-matter materials can be found, commercial use is unlikely, says Bausch.

**“We’re still kind of astonished at what can happen.”**

Advances in synthetic active materials may show the way forward. In 2013, New York University physicist Paul Chaikin and his colleagues described making particles of haematite, an iron oxide mineral, inside a spherical polymer<sup>7</sup>. When the scientists placed these ‘swimmers’ in a solution of hydrogen peroxide and exposed them to blue light, a chemical reaction caused the particles to move around spontaneously, clumping and unclumping like groups of people at a cocktail party.

In 2013, Bartolo and his colleagues reported large-scale flows using even simpler plastic spheres in a conducting fluid<sup>8</sup>. When the researchers turned on an electric field, the spheres began rotating in random directions. At high enough densities, interactions between nearby spheres caused them to spontaneously roll, flock-like, in the same direction.

Such lab-made materials remain primitive, however, compared with those produced in cells by 4 billion years of evolution. Dogic notes that the kinesins he uses are much more efficient than any human-made motors at converting energy to motion. And Bartolo is also quick to discourage talk of short-term pay-offs. “I’m not targeting a specific application,” he



## Smart swarms

A simple model of interactions among self-propelled particles can realistically simulate the movement of flocks of birds, schools of fish, self-assembling proteins in the cell and many other forms of active matter.

### Low density: randomness

When individuals have few neighbours to compare themselves to, they mill about with no obvious pattern.



### Higher density: flocking

As the density increases, the group's motion becomes synchronized.



says of his rotating plastic spheres.

Possible applications aside, active matter excites scientists because it so closely resembles the most complex self-organizing systems known: living organisms. In 2011, Dogic and his colleagues reported<sup>9</sup> that microtubule bundles anchored at one end to air bubbles on a microscope slide beat in synchronized, wave-like patterns eerily reminiscent of the hair-like cilia and flagella that protrude from the surfaces of some cells. And in his 2012 paper<sup>1</sup>, he noted a striking similarity between his microtubule flows and cytoplasmic streaming, a process in which cytoskeletal filaments team up to whisk a cell's contents around like "a giant washing machine", he says.

The resemblance between lab-prepared active matter and living things can be uncanny, agrees Jennifer Ross, a physicist at the University of Massachusetts Amherst. At talks, she has shown videos of spherical microtubule-kinesin systems and asked audience members whether they think they are seeing a real cell. "Whenever I present these to cell biologists in particular, they are always fooled," she says.

But something can look and act like a living organism without actually following the same rules, cautions Howard. He points out that Dogic's group created something that looks and acts very much like a cilium or flagellum with its multitude of proteins — but that may, in fact, work very differently. "There's something in there about the underlying mechanism, but it's extremely abstract," he says.

### IS IT ENOUGH?

To probe whether active-matter theory can reveal biological mechanisms, Daniel Needleman, a biophysicist at Harvard University in Cambridge, Massachusetts, studied the spindle: a microtubule-based structure that controls the separation of chromosomes during cell division. He wanted to test the idea, suggested by earlier theories and experiments, that short-range microtubule-kinesin interactions by themselves were sufficient to yield spindle-like structures. He first used sophisticated microscopes to examine extracts from frog egg cells, quantifying microtubule density, orientation and stresses during spindle formation. "It really was not clear at all until Dan came along that you could measure all these things," says Howard.

Needleman then merged his measurements with models of how active matter self-organizes. In 2014, he and Jan Brugués, a biologist at the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden, Germany, reported that, consistent with theory, the interactions they observed among closely spaced microtubules are enough to produce the spindle and keep it stable<sup>10</sup>. "People have argued that you need more complex processes," says Needleman. "But the fact that one can understand so much of the spindle without invoking any of that shows that it's certainly not necessary."

Others are using ideas from active matter to

probe how large numbers of cells organize in processes such as tissue growth, wound healing and the spread of tumours. Theorists including Marchetti, Joanny and Frank Jülicher of the Max Planck Institute for the Physics of Complex Systems in Dresden have modelled tissues<sup>11</sup> and tumours<sup>12</sup> as flowing cells that self-organize through short-range cell-to-cell interactions rather than chemical signals. Experimentalists are testing such ideas, for instance, by showing that active-matter theory can help to describe cell organization in a developing fruit-fly wing<sup>13</sup>.

Some biologists hope that such studies will reveal the fundamental principles that govern how cells divide, take shape or move. "It's like Linnaean classification before Darwin came along," says biologist Tony Hyman of the Max Planck Institute of Molecular Cell Biology and Genetics. "We've got all these molecules, just like they had all those species, and we need to put some kind of order, some kind of reason behind it all." Active matter, Hyman thinks, could provide that reason.

But even enthusiasts admit that mainstream biologists may need convincing. "We used to get a lot of papers rejected at the beginning," says Hyman — in part because the manuscripts' heavy use of mathematics made it hard to find reviewers. Even the phrase 'active matter' may hinder communication, adds Howard. "It's kind of a physics-y term."

Still, Howard and Hyman hope that acceptance will be aided by increasing convergence between fields. Among biologists, says Hyman, "I think the new generation coming along will be trained in physics from the beginning."

And that's good, adds Stephan Grill, a biophysicist at the Technical University of Dresden, because progress in active matter calls for scientists who are at the cutting edge of both physics and biology. "The pot of gold is at the interface," he says, "but you have to push both fields to their limits." ■

**Gabriel Popkin** is a freelance writer in Mount Rainier, Maryland.

1. Sanchez, T., Chen, D. T. N., DeCamp, S. J., Heymann, M. & Dogic, Z. *Nature* **491**, 431–434 (2012).
2. Nédélec, F. J., Surrey, T., Maggs, A. C. & Leibler, S. *Nature* **389**, 305–308 (1997).
3. Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I. & Shochet, O. *Phys. Rev. Lett.* **75**, 1226–1229 (1995).
4. Toner, J. & Tu, Y. *Phys. Rev. Lett.* **75**, 4326–4329 (1995).
5. Schaller, V., Weber, C., Semmrich, C., Frey, E. & Bausch, A. R. *Nature* **467**, 73–77 (2010).
6. Keber, F. C. et al. *Science* **345**, 1135–1139 (2014).
7. Palacci, J., Sacanna, S., Steinberg, A. P., Pine, D. J. & Chaikin, P. M. *Science* **339**, 936–940 (2013).
8. Bricard, A., Caussin, J.-B., Desreumaux, N., Dauchot, O. & Bartolo, D. *Nature* **503**, 95–98 (2013).
9. Sanchez, T., Welch, D., Nicastro, D. & Dogic, Z. *Science* **333**, 456–459 (2011).
10. Brugués, J. & Needleman, D. *Proc. Natl Acad. Sci. USA* **111**, 18496–18500 (2014).
11. Ranft, J. et al. *Proc. Natl Acad. Sci. USA* **107**, 20863–20868 (2010).
12. Basan, M., Risler, T., Joanny, J.-F., Sastre-Garau, X., & Prost, J. *HFSP J.* **3**, 265–272 (2009).
13. Aigouy, B. et al. *Cell* **142**, 773–786 (2010).

# COMMENT

**CULTURE** A year of science-inflected shows and events **p.22**



**LITERATURE** Looking forward to H. G. Wells celebrations **p.24**

**CLIMATE** Disjunct between Paris promises and investment at home **p.25**

**ORIGINS** Put disciplinary heads together on how life got started **p.25**

DENNIS SCHROEDER/NREL



Wind-turbine maintenance costs could be cut with a data-driven health-monitoring system.

## Share data on wind energy

Giving researchers access to information on turbine performance would allow wind farms to be optimized through data mining, says **Andrew Kusiak**.

The energy industry has long met demand by varying the rate at which it consumes fuel. Controlling the output of an oil-fired power plant is much like changing the speed of a car — press the accelerator pedal and more gas flows to the engine.

But the wind cannot be turned up or down. Smart software can make wind farms more efficient and responsive. Computer models can predict wind speed and control the number and capacity of turbines in operation to meet energy demand. Low-vibration designs and health monitoring would enable turbines to run more smoothly, avoiding expensive failures of gearboxes and other components whose replacement can cost hundreds of thousands of dollars and take days.

Optimizing renewables requires data: on device performance, energy output and weather predictions, seconds to days in advance. Vast quantities of information are collected by turbine manufacturers, operators and utility companies — yet hidden in their archives<sup>1</sup>. The information is prohibitively difficult for anyone outside to access.

It took me two years of discussions with different energy companies and the signing of several non-disclosure documents to obtain enough data to carry out a study on the performance of wind farms in Iowa, for instance. Wind-turbine data are usually recorded every 10 seconds and averaged over 10 minutes (see 'Poor performance'); getting higher-frequency data involves obtaining permissions from sensor manufacturers. Even basic data such as wind speeds and historical data on turbine operations were initially impossible to obtain. By approaching different partners and developing data-sharing agreements, we eventually gained limited access to wind energy data.

The lack of data sharing in the renewable-energy industry is hindering technical progress and squandering opportunities for improving the efficiency of energy markets. I call on the energy industry to follow the examples of defence, commerce and health care and share its data openly so that researchers can design better solutions for powering our planet.

There is money to be made. Academic ▶



► and industrial researchers need first to develop suitable wind-farm management models and prove their value. Software companies can sell energy and weather-monitoring and -predicting systems. Large technology companies such as the Hewlett Packard Enterprise or Google should establish wind-energy divisions for planning and balancing energy across different states and countries, as General Electric has done in wind-turbine manufacturing.

Leveraging renewable-energy data makes economic sense for a product — electricity — that is universal. Unlike other commercial industries, energy utilities do not compete on the basis of product quality but on generation and distribution processes and business operations, which are the greatest beneficiaries of big-data mining. Efficient renewable-energy plants equipped with software for accurate power prediction and responsive management will be able to take advantage of real-time, or 'spot', energy prices — supplying more when prices and demand are high and less when they are low. This extra profitability will encourage more firms and utility companies to acquire renewable-energy assets.

## DATA SCIENCE

The renewable-energy industry is awash with data. Wind-turbine manufacturers routinely collect data from hundreds of sensors on experimental and installed devices, measuring, for example, wind speed, oil temperature, vibration and power generation<sup>2</sup>. Utility companies record similar data from boilers and generators.

'Balancing authorities', usually non-profit, governmental or private organizations, match the expected demand for energy with the production scheduled by utility companies hours ahead of

generation. National, state and regional meteorological agencies and weather forecasters accrue radar data and run numerical weather-prediction models every 1–3 hours to produce forecasts and parameters such as wind speed.

New sources of data are emerging. The wind industry is experimenting with using sonar and laser-based lidar measurements to anticipate the speed, direction and turbulence of the wind approaching wind farms. Some utility companies fly drones over their farms to check turbine blades and measure wind speeds and directions to improve power prediction and to anticipate fluctuations over minutes to hours.

Renewable-energy producers operate in isolation. If industry players pooled their data and monitoring resources, they would all benefit. More-efficient and lower-cost wind-turbine designs could emerge, allowing turbines to last longer and produce more energy, and allowing output to be more accurately predicted. For example, combining data from wind farms in different US states would dramatically improve the accuracy of predicted hourly changes in power production.

Experiments that are impossible with a real wind turbine or a farm can be simulated on a digital replica<sup>3</sup>. Different control strategies can be tested for maximizing and smoothing the energy output. Conditions of components and subsystems could be analysed to lower maintenance costs — the most significant expense of wind-energy generation<sup>4</sup>. Active control of turbine vibrations could be studied. More stable turbines are less likely to fail and could be run beyond their current upper speed limit (usually around 20 metres per second) to produce more energy. The impact of atmospheric conditions on wind-farm sites

and energy production could be studied.

Controlling wind turbines with data-driven software could, models show, increase energy production by at least 10%, and gains of 14–16% are possible. Increasing the maximum running speed could easily add another 10%. Wind-farm maintenance costs could be cut by 10% with a data-driven health-monitoring system.

Yet the wind industry remains largely oblivious to data science<sup>5</sup>. A few utility companies are setting up in-house data-analytics teams, but the benefits of working

**"Defence, commerce and health-care organizations have developed processes for sharing data."**

with academic researchers and others are not recognized. Although models and software that do not directly impinge on turbine operations — such as a graphical display of

a turbine output — are broadly welcomed, direct interventions are impossible.

I have been unable to test control solutions developed in my Intelligent Systems Laboratory at the University of Iowa in any commercial settings. Even public utilities and colleges that own and operate wind turbines ended negotiations once they realized that their insurance and maintenance contracts would have to be modified. Wind-turbine insurance contracts tightly prescribe operating conditions and safety aspects, sometimes requiring turbines to be equipped with specific sensors (such as for tower vibration and rotor speed).

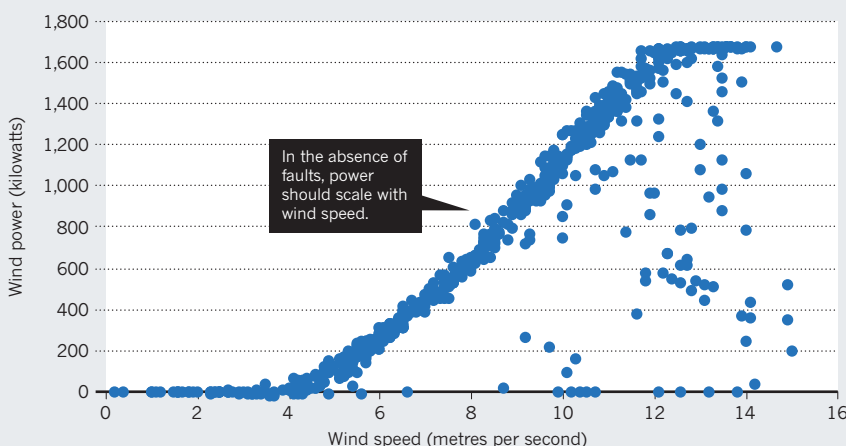
Potential for exposing flaws and poor design practices is another obstacle. Manufacturers may not want to reveal performance metrics that are covered by warranty terms or design details that might point to patent infringements. Competition is a worry.

## OPEN SHARING

Other sectors do better. Defence, commerce and health-care organizations have developed processes for sharing data with the research community while maintaining confidentiality and security. Some have created benchmark data sets to test data-analysis algorithms. Others run competitions to solve specific problems. For example, in 2006, the television- and film-streaming service Netflix offered a US\$1-million prize for an improved algorithm to predict rating scores of films. In 2011, the US National Renewable Energy Laboratory (NREL) ran the Round Robin project, in which they shared high-frequency vibration data from a healthy<sup>6</sup> and a faulty gearbox with competing teams to discover the most accurate ways to diagnose faults. It has been estimated that the

## POOR PERFORMANCE

Intermittent faults caused by blade misalignment or electronic problems, for example, can reduce the power produced by a single turbine. (Data taken at 10-minute intervals over 5 days.)





Turbine failures, resulting from gearbox and other component faults, could be avoided by sharing data.

value of the voluntary contributions to the project from the 16 participating teams (including the Intelligent Systems Laboratory at the University of Iowa) was worth between \$2 million and \$3 million.

Non-disclosure agreements outlining the specifics of data sharing and results dissemination are used in data-intensive projects. Consumer-goods company Procter & Gamble, for example, reveals information about a product (a new shampoo or a shaver, for instance) early in the design stage to potential customers, whose feedback improves the final design. On social-media platforms such as Facebook, users determine the scope of information sharing.

### NEW PROTOCOLS

The renewable-energy industry should adopt similar practices. First, it needs to decide which data can be shared and at what risk. Wind speed and direction, for example, could be released given that anyone could measure them. Although data on the real-time energy output of an entire wind farm should be rightfully protected for competitive reasons, sharing power produced by one or a few turbines would not compromise business value. When necessary, data could be transformed or anonymized; for example, by reporting relative percentage changes

rather than absolute power values.

Wind-energy associations in Asia, Europe, South America and North America should facilitate the data-sharing discussion. A summit of these players should define a path to open-access data as follows.

First, make all renewable-energy stakeholders aware of the problems and of the benefits of data sharing. Invite representatives from other manufacturing and service industries to present their data-sharing practices.

Second, develop data-sharing protocols and governance structures. US Department of Energy laboratories such as the NREL and Sandia National Laboratories could lead this effort because they collect renewable-energy data from some wind-farm operators for their own studies. Collecting data at higher frequencies (in some cases), at fraction-of-a-second intervals, from more utility companies and facilitating open access to them would be the next step. Although data collection should ideally be global, in reality, most useful results would be regional.

Third, develop a data-and-knowledge sharing platform for renewable energy. Stakeholders must decide how the data are to be assembled and pre-processed for use by the research community and industry.

Ideally, data would flow out to the research community and research results in the form of new models, algorithms, design solutions and other results would flow in. The vast majority of the results produced by the research community would remain open to review, scrutiny, future use and benchmark studies. Industry could retain ownership of the internally generated results as well as those produced by research contracts.

This long-awaited engagement will generate new science and greatly benefit renewable-energy companies, energy-equipment manufacturers and society by bringing more clean energy at a lower price. ■

**Andrew Kusiak** is professor of mechanical and industrial engineering, and director of the Intelligent Systems Laboratory, at the University of Iowa, Iowa City, Iowa, USA. e-mail: [andrew-kusiak@uiowa.edu](mailto:andrew-kusiak@uiowa.edu)

1. Kusiak, A., Verma, A. & Wei, X. *Wind Systems* **3**, 36–39 (2012).
2. Kusiak, A., Zhang, Z. & Xu, G. *IEEE Trans. Sustain. Energy* **4**, 756–764 (2013).
3. Zhang, Z., Zhou, Q. & Kusiak, A. *IEEE Trans. Sustain. Energy* **5**, 228–236 (2014).
4. Kusiak, A., Zhang, Z. & Verma, A. *Energy* **60**, 1–12 (2013).
5. Kusiak, A. *Ind. Eng.* **47**, 38–42 (2015).
6. Zhang, Z., Verma, A. & Kusiak, A. *IEEE Trans. Energy Convers.* **27**, 526–535 (2012).





Glacier National Park in Montana is one of more than 400 sites administered by the US National Park Service, which turns 100 this year.

#### LISTINGS

# Science in culture 2016

Gear up for some big birthdays, as anniversaries roll around for Star Trek, H. G. Wells and the US National Park Service. And jostling for the spotlight are *Finding Nemo*'s fishy crew, a modern twist on haute couture, groundbreaking artists, ground-quaking dinosaurs and (perhaps) Keanu Reeves. **Daniel Cressey** reports.

## 100 Years: The US National Park Service

On 25 August, the US National Park Service (NPS) celebrates its centenary — 100 years since President Woodrow Wilson signed it into existence, building on the 1872 creation of Yellowstone National Park by his predecessor Ulysses S. Grant. The service, which protects iconic landscapes from California's Yosemite to the Florida Everglades, is hosting events across its 409 sites. Of course, the greatest show of all, curated by nature itself, runs 365 days a year in NPS parks, from the Kilauea and Mauna Loa volcanoes on Hawaii's Big Island to the 3,516-kilometre Appalachian Trail on the US East Coast — the world's longest footpath open only to hikers.

## Dinosaur extravaganza

*American Museum of Natural History, New York City  
Starting in January*

New York gets a titanic new resident from 15 January, when the American Museum of Natural History installs a 37-metre-long cast of an as-yet-unnamed titanosaur. The bones of this giant herbivore, yet to be officially designated a species, were dug out of the Patagonian desert in Argentina. But that is just the start of the ancient-animal rollout. From March 2016 to January 2017, the exhibition *Dinosaurs Among Us* will explore how the titanosaur's relatives evolved to become birds: rare fossils and huge models will shed light on everything from birds' bones to dinosaur behaviour and brains. And from May 2016 to January 2017, *Ancient Predators in a Modern World* will tour 200 million years of crocodiles and their alligator, caiman and gharial relatives.

These remarkably specialized beasts are still recognizably the same as their ancestors that shared Earth with the dinosaurs.

## manus x machina

*The Metropolitan Museum of Art, New York City  
5 May – 14 August*

The Costume Institute at New York's Metropolitan Museum of Art aims to show the world that there is more to high fashion than pouts and peplums: technology and style have been in symbiosis from the off. From an 1880s Worth gown to a 2015 Chanel suit, this show contrasts and draws parallels between the handmade marvel of haute couture (*manus*) and machine-produced clothing (*machina*). Do technologies such as laser cutting, ultrasonic welding or 3D printing stand up against intricate embroidery and hand-stitched linings? Strut in for a look.

NPS/JACOB W. FRANK

### Moholy-Nagy: Future Present

Guggenheim Museum, New York  
June – September

This major retrospective comes 70 years after the death of Hungarian industrial designer and radical artist László Moholy-Nagy, whose oeuvre spanned photography, abstract painting and metal sculpture. The artist founded a school of design in Chicago, Illinois, and died in the city in 1946. He was a key player in the German Bauhaus movement, embracing the transformative power of technology and mechanization in kinetic artworks (like his contemporary, Alexander Calder; see [go.nature.com/n8rzsn](http://go.nature.com/n8rzsn)). In the early 1920s, Moholy-Nagy experimented with outsourcing paintings by describing them in detail over the telephone to a painter in a sign factory, with the aid of colour charts and graph paper. The exhibition travels to Chicago and Los Angeles, California, after its New York run.

### Finding Dory

Director: Andrew Stanton  
Opens 17 June

Digital-animation giant Pixar releases the much-anticipated follow-up to its 2003 *Finding Nemo*, a film so successful that clownfish are now often referred to as 'nemos'. The original had marine biologists in raptures over its faithfulness to the science. Pixar has a mixed record when it comes to sequels, but if *Finding Dory*, featuring Nemo's *Paracantharus* friend (pictured), can combine the remarkable accuracy with the superb storytelling that the company is capable of,

it could join Pixar's list of Oscar-botherers. Rumours suggest that the film was rewritten after the success of *Blackfish*, the 2013 documentary by director Gabriela Cowperthwaite that criticized the controversial keeping of killer whales in captivity.



### Engineering the World

Victoria and Albert Museum, London  
18 June – 6 November

Early in his career, engineer Ove Arup (1895–1988) worked on the floating Mulberry Harbours — temporary concrete breakwaters and piers set up for the Second World War Allied landing in Normandy on D-Day in 1944. He went on to help build iconic structures such as the Sydney Opera House, where his instincts for aesthetics and materials shone, before founding international mega-consultancy Arup, whose masterworks range from London's new research powerhouse, the Francis Crick Institute, to the Victoria and Albert Museum's own ongoing expansion plans. The exhibition promises to reveal Arup's multidisciplinary approach "as a humanistic and technological tool for social

### 50 Years: Star Trek

The world of Star Trek, first brought to television by US screenwriter Gene Roddenberry in 1966, inspires love in seemingly inverse proportion to the quality of its set design, special effects and (occasionally) acting. What keeps legion Trekkies passionate is the lingering glow of Roddenberry's delight in discovering "new life and new civilizations". Since the original show — with its then-radical multi-ethnic crew (some members pictured) — there have been another 4 live-action television series and 12 films, with fans from NASA leaders to schoolchildren. With the 13th film, *Star Trek Beyond*, due this year and a new television series promised in 2017, Star Trek remains the key science-fiction universe of modern times. (Sorry, Star Wars fans.)



PARAMOUNT TELEVISION/THE KOBAL COLLECTION

responsibility" and features prototypes, models and digital animations. It is part of the museum's Engineering the World exhibition, which will also include an installation by architect Achim Menges.

### Georgia O'Keeffe

Tate Modern, London  
6 July – 30 October

Visceral, often quasi-abstract evocations of botanical morphology in close-up paintings of irises and petunias helped to establish Georgia O'Keeffe's early career in 1920s New York. This retrospective at the Tate Modern will demonstrate the remarkable range of this groundbreaking modernist artist. In the late 1920s, O'Keeffe (pictured) moved figuratively and literally west to New Mexico, where the exposed, multicoloured topography inspired her to paint powerful landscapes and surreal studies that juxtaposed blossoms, bones, rocks and deadwood — paintings that in turn inspired others to re-examine the geology and biological riches of desert places.



### ANS (Autonomes NervenSystem)

Staatsoper, Berlin  
12–16 July

The autonomic nervous system controls unconscious human bodily functions such as heart rate and breathing. Greek-born musician Irini Amargianaki presents a multidimensional exploration of perception, incorporating instrumental pieces, video projections by Maryna Shuklina and shadow puppetry by Lisa Hauke. Premiering in Berlin as part of the Infecktion! festival for new musical theatre.

### Colour and Vision

Natural History Museum, London  
15 July – 6 November

How did colours appear in the living world, and how did animals evolve the ability to see them? This exhibition at London's Natural History Museum will take viewers from the eyes of the beholder to the art and innovation that have emerged from nature's wild palette. Colour is crucial — from lights telling you to go or stop to tropical frogs that sport pigments screaming 'poison'.

### The Universe and Art

Mori Art Museum, Tokyo  
30 July – 9 January 2017

This exhibition asks how humans have viewed the Universe through millennia, starting with *The Tale of the Bamboo Cutter* — the oldest known piece of narrative prose in Japan, which dates back to the tenth century — and zipping forward to the age of the International Space Station and the search for alien life. Ranging from ancient art to contemporary astronomy, the show promises to explore how people through the ages have conceived of the vastness around them.

TONY VACCARO/GETTY





H. G. Wells (right) wrote more than 100 books, including science-fiction classic *The War of the Worlds* (top).

#### 150 Years: H. G. Wells

This year sees two key dates focused on prescient author, scientist and educator H. G. Wells: the 150th anniversary of his birth, and the 70th of his death. Wells, who wrote more than 100 books, including *The Time Machine* (1895), *The War of the Worlds* (1898) and *The Island of Doctor Moreau* (1896), transformed turn-of-the-century science into literature that is still read, dissected and argued over today. Some of his short stories will be brought to life in a series of dramas airing on UK channel Sky Arts, and the H.G. Wells Society plans a programme of events (see [go.nature.com/aasjbm](http://go.nature.com/aasjbm)).



#### ▶ Deepwater Horizon

Director: Peter Berg  
Opens 30 September

The 2010 Deepwater Horizon explosion and oil spill — which killed 11 people and spewed more than 3 billion barrels of crude oil into the Gulf of Mexico — was the worst environmental disaster in recent US history. Ecosystems and the regional economy suffered: wildlife died, beaches were shut and fisheries floundered. Now the events leading up to the disaster get the Hollywood treatment, as Mark Wahlberg takes on the role of an electronics technician on the doomed drilling rig. Expect gritty drama and courage in the face of adversity.

#### Replicas

Director: Tanya Wexler

Few details were available on this science-fiction thriller as *Nature* went to press — not normally a good sign in a film. But reports that Keanu Reeves will play a neuroscientist who fights the government, police and the very laws of science to resurrect his family are intriguing.

## SECOND CHANCES

IF YOU MISSED THESE SHOWS THE FIRST TIME ROUND, THEY ARE NOW ON TOUR

#### The Douanier Rousseau — Archaic Candour

Musée d'Orsay, Paris  
22 March – 17 July

Dozens of pieces by French post-Impressionist Henri Rousseau go on display at the Musée d'Orsay in Paris after a stint in Italy. They include stunning depictions of jungles, created by a man who never left France but regularly visited the botanical gardens of Paris.

#### Leonardo da Vinci: The Mechanics of Genius

Science Museum, London  
10 February – 4 September

Who needs the *Mona Lisa* when you can have a flying machine? This touring exhibition of ingenious models brings to life the mechanical contraptions sketched out by history's greatest polymath.

#### Strandbeest: The Dream Machines of Theo Jansen

Exploratorium, San Francisco, California  
27 May – 5 September

Arriving in California in May, far from their birthplace on the Dutch coast, are the vast, surreal-looking Strandbeest ('beach beast') automata created by artist Theo Jansen. These mechanical animals move on gusts of wind and have amazed viewers with their strange dances, at once robotic and naturalistic. ■

BETTMANN/CORBIS

AF ARCHIVE/ALAMY

# Correspondence

## China boom leaves children behind

Some 61 million children were left behind by migrant parents in China in 2010–14, or almost 22% of the country's children (see [go.nature.com/clylrn](http://go.nature.com/clylrn); in Chinese). This side effect of urban development is seriously affecting the mental and physical health of those abandoned juveniles who are uncared for. We urge China's government to weigh this social damage against economic gain and to take steps to mitigate it.

There were 274 million Chinese migrant workers in 2014 — an unprecedented number. Evidence is increasing for the adverse effects of such upheaval on some children's physical, psychological and social development (Q. Li *et al. China Econ. Review* **36**, 367–376 (2015); G. Ding and Y. Bao *J. Child Psychol. Psychiatr.* **55**, 411–412 (2014)).

Despite China's great economic achievements, its childcare services remain underfunded and underdeveloped. There is an urgent need for policy reform and strategies to tackle the problem. These include the development and enforcement of family interventions, community support and schooling improvements.  
**Peng Yuan, Long Wang**  
*Xiangya Hospital, Central South University, Changsha, Hunan, China.*  
[wanglong@csu.edu.cn](mailto:wanglong@csu.edu.cn)

## Recover wastewater resources locally

As contributors to governmental initiatives to reuse wastewater pollutants in the European Union, the United States and China, we consider that decentralized recovery of these resources could result in more environmental, economic and social benefits than the near-term upgrade of centralized facilities

(see W.-W. Li *et al. Nature* **528**, 29–31; 2015).

Decentralized local treatment and reuse facilities avoid the large transportation and energy costs of conveying treated wastewater back to catchment areas for reuse. Concentrating nutrients for recovery also consumes large amounts of energy: urine makes up only 1% of total wastewater volume, and about 80% of nitrogen and 50% of phosphate in wastewater are from urine.

However, nitrogen and phosphorus can be locally recovered from urine using urine-diverting toilets, which greatly reduce nutrient loads to existing plants (J. Elser and E. Bennett *Nature* **478**, 29–31; 2011). Similar strategies can be applied to carbon, leading to greater energy recovery through co-digestion of sludge with food waste, and direct carbon capture and storage for climate-change mitigation.

International strategies for nutrients, energy and water ('NEW' initiatives) aim to transform the water infrastructure for resource recovery. By balancing near-term goals and long-term ambitions, water 'waste' should become a misnomer.

**Zhiyong Jason Ren** *University of Colorado, Boulder, USA.*  
**Art K. Umble** *MWH Global, Denver, USA.*  
[zhiyong.ren@colorado.edu](mailto:zhiyong.ren@colorado.edu)

## Half of samples fail protein-blot tests

Poorly characterized antibodies give rise to irreproducible results (see, for example, *Nature* **527**, 545–551; 2015), but so can the use of properly validated antibodies in a non-validated context.

At Aviva Systems Biology in California, we used our highly specific commercial antibodies in western immunoblot assays to test more than 1,000 protein samples provided by the research community. We found that the

preparation quality of more than half of these samples failed to meet the technical requirements for a reliable assay signal.

Simple technical factors confounded the electrophoretic resolution or antibody detectability of the researchers' protein solutions. These included unsuitable sample concentrations, buffer incompatibility and the absence of calibration markers or treatment controls. Until uniform western-blotting standards are widely adopted (see J. E. Gilda *et al. PLoS ONE* **10**, e0135392; 2015), there is a risk that data irreproducibility will continue to be the norm.

Antibody-production companies should not be treated as casinos for boosting a researcher's chances of a positive result from such shot-in-the-dark samples.

**Matt Landry** *Aviva Systems Biology, San Diego, California, USA.*

**Aldrin V. Gomes** *University of California, Davis, USA.*  
[mlandry@avivasysbio.com](mailto:mlandry@avivasysbio.com)

## UK budget cuts erode Paris promises

Two weeks before the UK government signed up to keep global warming well below 2 °C at the 2015 United Nations climate summit in Paris, it announced a 22% budget cut for the Department of Energy and Climate Change. It also scrapped a previously ring-fenced £1-billion (US\$1.5-billion) budget for developing carbon capture and storage. UK decarbonization targets might be unachievable without this technology.

In our view, these actions signal that the UK government does not treat climate action as a priority, and that it is ignoring the evidence of the research it funds. For example, data collected by the UK Met Office show that 2014 and 2015 have been the warmest years on record.

To regain credibility, the government must overcome internal division (N. Carter and B. Clements *Br. Politics* **10**, 204–225; 2015) and develop a robust climate policy that is in line with its stated ambitions.  
**Alexander C. Lees** *Cornell University, Ithaca, New York, USA.*  
**Andrew Balmford, Ben Phalan** *University of Cambridge, UK.*  
[bt22@cam.ac.uk](mailto:bt22@cam.ac.uk)

## Come together to study life's origins

Researchers working on the origins of life tend to fall into two camps — those who investigate artificial life and those who study the origins of life on Earth four billion years ago. The communication gulf between the two needs to be closed if the field is to progress.

Artificial-life researchers are less concerned about how life originated on Earth than with the idea of life as a universal phenomenon — including its emergence and self-organization. And those pursuing experimental verification of mechanisms for terrestrial origins are seldom drawn to the broad theoretical ideas of artificial life.

The Earth-Life Science Institute's Origins Network, working with members of the research community, has issued a statement to encourage fresh approaches to the subject (C. Scharf *et al. Astrobiology* **15**, 1031–1042; 2015). We suggest that origins-of-life research requires inspirational innovation, cross-disciplinary collaboration and reassurance from institutions that such research will be supported. We hope that these proposals will help to train a new generation of scientists to think more broadly and less tribally.  
**Caleb Scharf** *Columbia University, New York, USA.*  
**Nathaniel Virgo, H. James Cleaves** *Earth-Life Science Institute, Tokyo, Japan.*  
[caleb@astro.columbia.edu](mailto:caleb@astro.columbia.edu)



## ASTROPHYSICS

# Why black holes pulse brightly

Black holes can produce oscillating outbursts of radiation that were thought to be associated with high rates of infalling matter. The observation of pulses of visible light from a black hole complicates this picture. [SEE LETTER P.54](#)

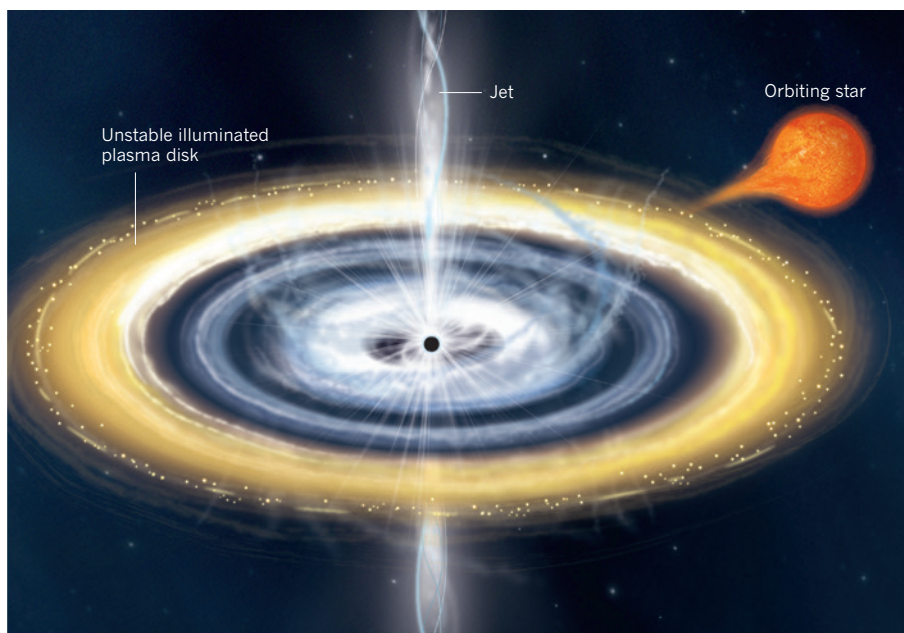
POSHAK GANDHI

Accretion of matter onto black holes is an efficient way of converting mass into energy, much more so than the process of nuclear fusion, which powers the light from stars. But unlike fusion, the physics behind accretion is still not understood, more than 40 years after the identification of accreting black holes in the Milky Way<sup>1,2</sup>. On page 54 of this issue, Kimura *et al.*<sup>3</sup> present exquisite observations made during a black-hole accretion episode. They show that the visible radiation from the black hole's vicinity oscillates dramatically — sometimes regularly, other times not — in a manner not predicted by models. Such oscillations were previously associated with high rates of infalling matter, but the authors report that the observed oscillations can occur even when the rate of infall is low. Understanding this behaviour could help astronomers to better understand violent accretion episodes onto black holes.

The researchers studied the black hole V404 Cygni, which is 2.4 kiloparsecs from Earth. The Cygnus constellation is a popular area of the sky for black-hole specialists because it hosts several other bright, accreting black holes and neutron stars. In June 2015, V404 Cygni underwent a short-lived accretion 'outburst' that lasted for about two weeks, causing it briefly to become one of the brightest cosmic X-ray sources beyond the Solar System. The black hole's gravity is strong enough to strip matter off the surface of an orbiting companion star, and the potential energy of this infalling matter is released, in part, as the observed electromagnetic radiation.

The infalling matter is thought to be hot, magnetized plasma. But if this material were to plunge directly into the black hole, its energy would be lost immediately without any brightening. The standard picture of accretion is that the plasma instead acts as a viscous fluid that spirals towards the black hole in the form of a disk, and that its energy is liberated as a result of friction in the disk. Any plasma that cannot be accreted is expelled in the form of a fast narrow stream (a jet) or as an outflowing wind.

If there is a balance between the accreting plasma and frictional energy losses, then the mass is steadily accreted. But naturally



**Figure 1 | Light pulses from irregular accretion onto black holes.** Black holes can accrete matter from orbiting stars. The matter is thought to spiral towards the black hole as a plasma disk, and friction in the disk causes energy to be released in the form of electromagnetic radiation. Plasma that cannot be accreted is expelled as a jet. Kimura *et al.*<sup>3</sup> propose that, in systems such as V404 Cygni, the supply of infalling matter cannot steadily fill the disk between the companion star and the black hole, causing fluctuations in the density of matter in the disk. These fluctuations trigger oscillating emissions of X-rays (white lines radiating from black hole) near the black hole that ionize hydrogen atoms in the outer part of the disk and cause pulses of visible light (gold region), as observed by the authors.

occurring changes in the rate of mass accretion can upset this balance and cause an unstable see-saw-like behaviour: periods of enhanced accretion that empty parts of the disk are followed by quieter periods when the parts are refilled, after which the cycle begins again. An approximate analogy is the repetitive filling and emptying of a Japanese bamboo fountain.

Such behaviour has been observed in one other black-hole system, GRS 1915+105 in the Aquila constellation, which undergoes high levels of mass accretion. Several classes of repetitive oscillation occur in this system, but only in its observed X-ray emission<sup>4</sup>. Kimura and collaborators draw parallels between GRS 1915+105 and the visible-light oscillations in V404 Cygni, but make the crucial distinction that the latter oscillations occur at a much lower rate of mass accretion than the former ones. In other words, the repetitive

behaviour is not strictly associated with episodes of high mass accretion.

V404 Cygni is an important study target for several reasons. It was the first Galactic object to have its mass (nine solar masses) firmly placed within the range of masses associated with black holes<sup>5,6</sup>. Its distance from Earth is also known with higher accuracy than those of other black holes<sup>7</sup>. Moreover, it looks extremely bright when it accretes matter, despite being partly veiled behind interstellar gas and dust. In the absence of this veil, V404 Cygni would have been one of the most distant objects in the Milky Way visible in dark skies to the unaided eye in June 2015. Because V404 Cygni is so well characterized, Kimura *et al.* are able to propose a mechanism to explain the visible-light oscillations.

The authors suggest that in systems such as V404 Cygni and GRS 1915+105 there is a

relatively large volume of space between the companion star and the black hole, which allows a large disk to form. But the supply of infalling matter from the companion star is insufficient to fill such a large disk with a steady flow. Without a steady flow, the accretion rate becomes unstable and can fluctuate violently (Fig. 1). These fluctuations, in turn, trigger oscillating emissions of energetic X-ray photons near the black hole, which then light up the whole disk with the observed pulsating visible effects.

But the authors show that this explanation requires the disk to be very large, close to its maximum possible size. Moreover, the X-ray oscillations that they observed from V404 Cygni are much stronger than the visible-light ones. These puzzling facts will need to be accounted for. How, and whether, the jet of the black hole tracks these oscillations is also yet to be determined. The proposed parallels

between the observed oscillations and those of GRS 1915+105 will undoubtedly be investigated in detail in the future. This will help researchers to understand the above issues in light of the wealth of supporting observations currently being analysed by astronomers the world over.

Black-hole outbursts are unpredictable and some can be two weeks or even shorter in duration, so worldwide coordination and round-the-clock monitoring is essential if we are to understand the physics of these extreme events. This becomes particularly challenging when coordinating observations between space telescopes and those on the ground. The outburst of V404 Cygni last year invigorated the efforts of black-hole astronomers to tackle these challenges, with at least one conference dedicated entirely to this theme. Amateurs can also play a key part in this effort. Kimura and colleagues gathered data from many small

telescopes, some with optical elements only 20 centimetres in diameter, showing that, in astronomy, size is not necessarily what matters; collaboration does. ■

**Poshak Gandhi** is in the Department of Physics & Astronomy, University of Southampton, Highfield, Southampton SO17 1BJ, UK.

e-mail: p.gandhi@soton.ac.uk

1. Bowyer, S., Byram, E. T., Chubb, T. A. & Friedman, H. *Science* **147**, 394–398 (1965).
2. Webster, B. L. & Murdin, P. *Nature* **235**, 37 (1972).
3. Kimura, M. *et al. Nature* **529**, 54–58 (2015)
4. Fender, R. P. & Belloni, T. *Annu. Rev. Astron. Astrophys.* **42**, 317 (2004).
5. Casares, J., Charles, P. A. & Naylor, T. *Nature* **355**, 614–617 (1992).
6. Khargharia, J., Froning, C. S. & Robinson, E. L. *Astrophys. J.* **716**, 1105 (2010).
7. Miller-Jones, J. C. A. *et al. Astrophys. J.* **706**, L230–L234 (2009).

## ECOLOGY

# Different worlds

**Patterns of species association reveal that terrestrial plant and animal communities today are structured differently from communities spanning the 300 million years that preceded large-scale human activity. SEE LETTER P.80**

GREGORY P. DIETL

The British author L. P. Hartley wrote in one of his best-known novels, *The Go-Between*, that “The past is a foreign country: they do things differently there.” This poignant imagery of remoteness from the past captures the essence of an emerging global consciousness. Human hegemony over nature has become so pervasive and profound that it is quite possible that we have created a world that has little or no precedent — in ecological parlance, it has no analogue. On page 80 of this issue, Lyons *et al.*<sup>1</sup> detail a compelling case that this extraordinary situation

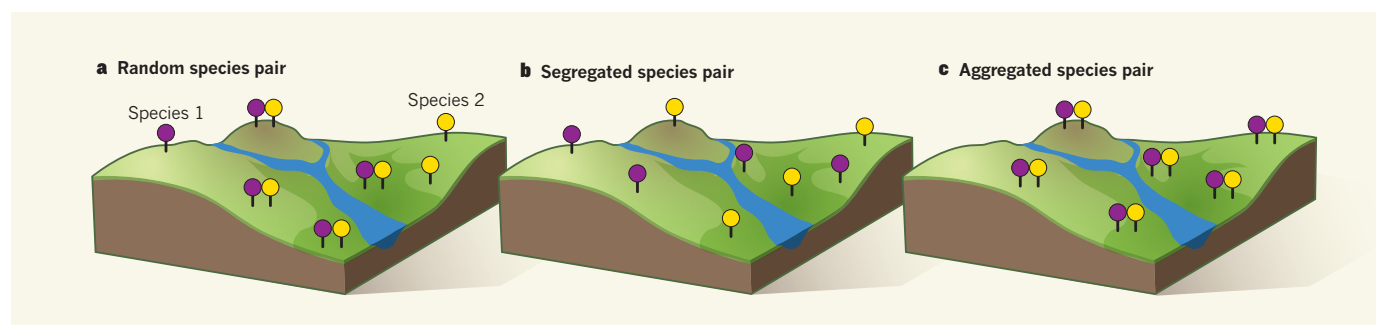
is an undeniable reality for the rules that govern how plant and animal communities are structured.

The authors assembled data on the presence and absence of terrestrial plant and animal taxa for 80 fossil and modern assemblages in North America, Africa and Eurasia, spanning the past 300 million years. Using a statistical approach that was designed to compare occurrence data against a randomized ‘null’ assemblage, they quantified the fraction of species pairs in each assemblage that deviated from random expectations about where they should be found. Species pairs meeting this criterion provide valuable

insight into the ecological processes that structure communities<sup>2,3</sup>.

In modern communities, most species pairs show random co-occurrence, but those that are non-randomly associated are typically segregated — that is, they tend to co-occur less frequently than would be expected by chance<sup>2</sup> (Fig. 1). Lyons *et al.* wanted to know whether the fossil record is consistent with this pattern of species segregation. The headline finding is that the pattern of co-occurrence dominating modern communities departs sharply from that of the past. As in modern communities, co-occurrence of most pairs was random. But unlike in modern communities, the non-random associations were dominated by aggregated species pairs, which co-occur more frequently than would be expected by chance<sup>2</sup> (Fig. 1). This dominance of aggregated pairs persisted with little change for more than 300 million years on different continents and across diverse taxa, until about 6,000 years ago, when the sharp transition to the segregated co-occurrence pattern began.

After running a battery of tests to ascertain that this temporal trend was not an artefact,



**Figure 1 | Species associations.** In a terrestrial ecological community, any two species may occur randomly (a) at locations in a landscape. Alternatively, species pairs may be non-randomly associated, in which case they can be either segregated (b), meaning that the two species co-occur less frequently than would be expected by chance, or aggregated (c), meaning

that they co-occur more frequently than expected by chance. Among non-randomly associated pairs, Lyons *et al.*<sup>1</sup> document a shift from a dominance of aggregated pairs before the expansion of human populations to the segregated pattern typically seen today. (Figure adapted from ref. 7.)



the authors speculate that an expanding human population may explain why species co-occurrence patterns are so different today. The shift was most obvious in North American assemblages (where the most occurrence data were available) and coincided with the inexorable spread of agriculture in this region. The authors propose that habitat fragmentation and limitations on species dispersal associated with land use were probably the main engines driving the shift. The structure of plant and animal terrestrial communities would never be the same again.

This interpretation is sure to attract fervent debate and lead to further research to confirm the pattern and disentangle the proposed mechanisms involved. The tension between the distant past and the familiar present that the study highlights, however, has an underlying implication that may not be as obvious. If the past is different from the present (in this case not in the imminent processes that were operating, but in their frequency), its applicability to our current societal need to anticipate ecological changes and design adaptation measures — a goal that Lyons *et al.* acknowledge is a priority — is not immediately manifest. There is no easy way around this tension. At stake is whether we can reliably use the past as a guide to an uncertain, anthropogenically modified future.

A small cadre of voices argues that a human-dominated present limits the use of the past as a key to unlocking the future<sup>4</sup>. In this view, the world we live in today, and the immediate future that our grandchildren will inherit, has no analogue in the geological past. As a consequence, referencing 'natural experiments' in the distant past as a guide to predict what might happen, now or in the future, is a flawed strategy. Out is the use of uniformitarianism<sup>5</sup> as a guiding principle, and in is a new kind of 'post-normal' science<sup>6</sup>. Lyons and colleagues' study of human impacts on community-assembly rules, at least as implicated by species co-occurrence patterns, seems to embody evidence for this no-analogue world.

A more optimistic view of this tension between the past and present — one that acknowledges that processes change and interact in complex ways over time, whether human action is involved or not — is that it poses a challenge for how we select analogues from the past to gain insight into future conditions. Lyons and colleagues' finding is a stark reminder that analogue selection often over-stresses likenesses at the expense of differences. However, small and unknowable differences in starting points may overwhelm the signal of the likenesses, making analogue selection a risky business. To use the past as a guide, we must select from the dense fabric of likenesses and differences that was its contingent state at a moment in time, and apply only those particular events and conditions relevant to our present needs.

Moving beyond this tension will require creative ways of thinking about how we use

the distant past to improve our understanding of the present and our anticipation of the future, which may provide a ground for wiser action. Lyons and colleagues' study is an excellent entry point into thinking about this problem. ■

**Gregory P. Dietl** is at the *Paleontological Research Institution, Ithaca, New York 14850, USA* and in the *Department of Earth and Atmospheric Sciences, Cornell University, Ithaca*.

## VIROLOGY

# Host protein clips bird flu's wings in mammals

**The polymerase enzyme from avian influenza A viruses does not function well in human cells. The protein ANP32A has been identified as the cellular factor mediating a major component of this host restriction. [SEE LETTER P.101](#)**

ANICE C. LOWEN

Influenza A viruses circulate in diverse natural hosts, including mammalian and avian species. Yet transmission of these viruses between mammals and birds occurs only rarely, owing to host restriction: an influenza A virus that is adapted to an avian host typically does not grow well in a mammalian host, and vice versa. When such restrictions are overcome and an avian virus transmits to humans, a pandemic can occur. On page 101 of this issue, Long *et al.*<sup>1</sup> report a breakthrough in understanding the restriction of avian influenza viruses in mammals.

The protein PB2 is a necessary component of the influenza A polymerase enzyme complex, which copies the viral genome and thus is essential for viral replication. For many years, researchers have known that a specific domain of PB2, the 627 domain, is involved in host restriction<sup>2</sup>. H5N1 strains and other 'bird flu' viruses rapidly acquire mutations in this domain following transmission to humans or inoculation of mammals in the laboratory<sup>3,4</sup>. These mutations, in turn, greatly enhance the growth, virulence and transmission of avian influenza A viruses in mammals<sup>5–8</sup>. Yet despite intense effort, the host factors and mechanisms that limit the functionality of non-mutated avian-adapted PB2 proteins in mammalian cells<sup>9–14</sup> remained obscure.

Long *et al.* knew from previous work<sup>15</sup> that the avian-adapted PB2 did not work well in mammalian cells because of the absence of a factor that enhances polymerase activity in avian cells, rather than because of the presence of an inhibitory factor in mammalian

*e-mail: gpd3@cornell.edu*

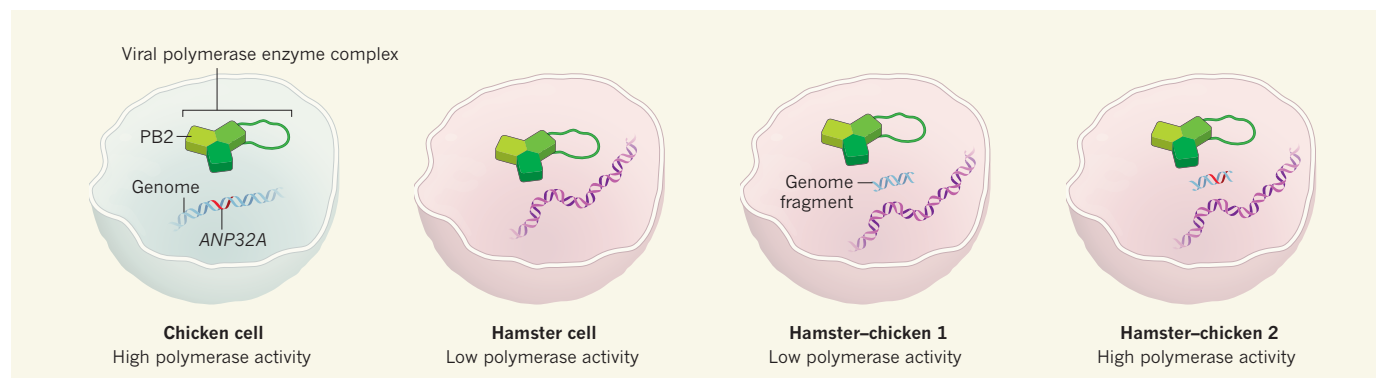
1. Lyons, S. K. *et al.* *Nature* **529**, 80–83 (2015).
2. Gotelli, N. J. & Ulrich, W. *Oecologia* **162**, 463–477 (2010).
3. Connor, E. F., Collins, M. D. & Simberloff, D. *Ecology* **94**, 2403–2414 (2013).
4. Knight, J. & Harrison, S. *Anthropocene* **5**, 71–75 (2014).
5. Baker, V. R. *Anthropocene* **5**, 76–79 (2014).
6. Funtowicz, S. O. & Ravetz, J. R. *Futures* **25**, 739–755 (1993).
7. Blois, J. L. *et al.* *Ecography* **37**, 1095–1108 (2014).

This article was published online on 16 December 2015.

cells. To identify the missing positive factor, the authors used a panel of hybrid hamster cell lines that each carried a different fragment of the chicken genome. They expressed an avian-adapted PB2 protein, along with its essential viral partner proteins, in each cell line and measured the activity of the viral polymerase (Fig. 1). Out of 53 hybrid cell lines tested, four showed robust activity of the avian-adapted polymerase complex. By identifying the chicken genes that were shared by these four cell lines, Long *et al.* narrowed their search for the positive avian factor to just 12 genes. Then, by expressing each of the candidate genes singly in mammalian cells, the authors found what they were looking for: chicken ANP32A is a single gene that enables an avian-adapted PB2 protein to function efficiently in mammalian cells.

Confirmation that ANP32A protein supports influenza-polymerase activity was obtained by decreasing the expression of ANP32A in cells. When levels were reduced in chicken cells, the activity of an avian-adapted viral polymerase decreased. Similarly, when expression of the mammalian version of ANP32A was reduced in human cells, a human-adapted viral polymerase was less active. Thus, ANP32A is crucial for influenza A virus replication in both birds and mammals, but avian-adapted polymerases work inefficiently with mammalian ANP32A. These findings indicate that the adaptive changes that influenza viruses acquire in the PB2 627 domain following transmission to mammals allow the viral polymerase to partner with mammalian ANP32A.

The researchers report that chicken and



**Figure 1 | A cellular factor for host restriction.** The influenza virus's polymerase enzyme complex is essential for viral replication in a host cell. One component of this complex, the protein PB2, contributes to host restriction — the fact that avian-adapted viral polymerases do not function properly in mammalian cells. To identify the host protein that partners with PB2 to cause restriction, Long *et al.*<sup>1</sup> used a panel of hybrid

hamster–chicken cell lines, each containing a different region of the chicken genome. They expressed avian-adapted influenza polymerase in these cells and measured its activity; by comparing the chicken genomic components of cell lines with polymerase activity, the authors identified the host gene *ANP32A* as underlying host restriction of viral activity.

human ANP32A proteins are similar except for a stretch of 33 amino acids that is missing from the human protein. All avian ANP32A genes, except those of ostriches and other ratites, encode these 33 amino acids, whereas all mammalian versions lack this region. Fittingly, addition of this sequence to a mammalian ANP32A protein was sufficient to permit avian-influenza PB2 function in mammalian cells. With this finding, what is known of bird flu in ostriches now makes perfect sense: influenza viruses isolated from ostriches tend to carry a PB2 with a mammalian-like sequence in the 627 domain<sup>16</sup>.

Long *et al.* have identified a host-cell protein that has an important function in the life cycle of influenza A viruses and that is a major factor in their host specificity. But it is still unclear how the virus uses ANP32A. The authors show that the protein does not alter the expression of PB2 nor its accumulation in the cell nucleus, where the viral genome is replicated. Is it instead directly involved in

the replication of viral RNA? Relatively little is known about the host requirements for this step in the life cycle of the virus.

Investigating the precise relationship between PB2 and ANP32A will not only give insight into the mechanism of influenza host restriction, but may also trigger further discovery of virus–host interactions that contribute to viral RNA replication. Moreover, the influence of adaptation in the PB2 627 domain on viral fitness suggests that disrupting the virus–ANP32A interaction could be a powerful means of controlling influenza infection. Therefore, elucidation of ANP32A's role in virus replication in molecular detail may open the way to the development of new antiviral drugs. ■

**Anice C. Lowen** is in the Department of Microbiology and Immunology, Emory University School of Medicine, Atlanta, Georgia 30322, USA.  
e-mail: anice.lowen@emory.edu

- Long, J. S. *et al.* *Nature* **529**, 101–104 (2015).
- Subbarao, E. K., London, W. & Murphy, B. R. *J. Virol.* **67**, 1761–1764 (1993).
- de Jong, M. D. *et al.* *Nature Med.* **12**, 1203–1207 (2006).
- Danzon, S. *et al.* *J. Virol.* **88**, 13436–13446 (2014).
- Steel, J., Lowen, A. C., Mubareka, S. & Palese, P. *PLoS Pathog.* **5**, e1000252 (2009).
- Munster, V. J. *et al.* *J. Infect. Dis.* **196**, 258–265 (2007).
- Hatta, M., Gao, P., Halfmann, P. & Kawaoka, Y. *Science* **293**, 1840–1842 (2001).
- Linster, M. *et al.* *Cell* **157**, 329–339 (2014).
- Mehle, A. & Doudna, J. A. *Cell Host Microbe* **4**, 111–122 (2008).
- Gabriel, G., Herwig, A. & Klenk, H. D. *PLoS Pathog.* **4**, e11 (2008).
- Paterson, D., te Velthuis, A. J., Vreede, F. T. & Fodor, E. *J. Virol.* **88**, 339–344 (2014).
- Weber, M. *et al.* *Cell Host Microbe* **17**, 309–319 (2015).
- Rameix-Welti, M. A., Tomoiu, A., Dos Santos Afonso, E., van der Werf, S. & Naffakh, N. *J. Virol.* **83**, 1320–1331 (2009).
- Mänz, B., Brunotte, L., Reuther, P. & Schwemmler, M. *Nature Commun.* **3**, 802 (2012).
- Moncorgé, O., Mura, M. & Barclay, W. S. *J. Virol.* **84**, 9978–9986 (2010).
- Shinya, K. *et al.* *J. Virol.* **83**, 13015–13018 (2009).

## ARCHAEOLOGY

# Sources of Chaco wood

**Tree rings can pinpoint the source of wood as well as how old it is. This method has now been used to identify the sources of timber used by the Native Americans who constructed the pre-Columbian 'great houses' of Chaco Canyon.**

**JARED DIAMOND**

The largest buildings erected by Native Americans in North America before European arrival were those at Chaco Canyon in what is now New Mexico, where the Ancestral Puebloan culture thrived between about AD 850 and 1140 (refs 1, 2). One of the unanswered questions that those

buildings pose is: where did the Ancestral Puebloans obtain large logs for their buildings in a desert with few trees? Writing in *Proceedings of the National Academy of Sciences*, Guiterman *et al.*<sup>3</sup> have used the method of tree-ring sourcing — not dating, but sourcing — to identify the origin as mountain forests more than 75 kilometres away and 1,000 metres higher up.

Modern visitors to Chaco are astonished that Ancestral Puebloan agriculture supported a complex society of thousands of people in a fragile dry environment where few try to farm today. Yet this culture succeeded for centuries, using sophisticated methods of managing water run-off from brief downpours. Their largest buildings, termed 'great houses', rose up to six storeys and contained hundreds of rooms; they remained North America's tallest buildings until steel construction finally permitted them to be topped by Chicago skyscrapers in 1885.

Great houses were built primarily of sandstone masonry, but they depended on wood for the beams of roofs, doors and windows (Fig. 1). Around a dozen great houses were constructed over three centuries, which required huge quantities of wood: about 240,000 trees were used, yielding beams up to 5 metres long and





**Figure 1 | Great houses.** **a**, The monumental architecture of the Ancestral Puebloan culture, from between around AD 850 and 1140, is evident at the ruins in Chaco Canyon, New Mexico. **b**, Built of sandstone masonry, the 'great houses' incorporated large wooden beams for roofs, doors and windows, but the desert environment lacked large trees. Guiterman *et al.*<sup>3</sup> used tree-ring-sourcing methods to identify the distant mountain ranges from which this wood was obtained.

weighing up to 300 kilograms<sup>4</sup>. But big trees are not abundant near Chaco today, and were probably not in Ancestral Puebloan times either. Furthermore, some trees used in the construction were spruce and fir, which grow only in mountain forests at elevations much higher than Chaco.

Previous studies had identified the wood as coming from the Chuska Mountains and Mount Taylor, which are both more than 75 km in a straight line from Chaco. These origins were determined by comparing the ratios of the strontium isotopes <sup>86</sup>Sr and <sup>87</sup>Sr in great-house beams with the ratios in trees from local mountain ranges<sup>5,6</sup>. (The ratio varies between trees from different mountains, depending on differences in the age and mineral content of the underlying rock.) The studies assumed from local palaeoecology that the Chaco Basin itself was unforested during Ancestral Puebloan occupation. But the isotope-analysis results and that assumption were subsequently criticized<sup>7</sup>.

Guiterman *et al.* turned to tree-ring sourcing for further evidence. In areas with strongly seasonal climates, tree wood displays annual growth rings. Tree growth, and hence ring thickness, differs from year to year because of annual differences in temperature and rainfall. This process underlies the familiar method of dendrochronology, better known as tree-ring dating: comparing the tree-ring pattern in an archaeological wood sample with the pattern in a tree sampled in a known year. But tree growth can also vary locally in a given year, because of local differences in climate and topography. That fact enables tree rings to be used to identify the source of an archaeological wood sample, by controlling for date and then comparing the sample's tree-ring pattern with patterns of wood from different local sources. Tree-ring sourcing has been used in Europe to identify sources of wood for ships, musical

instruments and paintings<sup>8</sup>, but the method has received less attention in North America.

The authors assembled tree-ring patterns from eight mountains located in a circle around Chaco. The patterns at the sites differed enough for wood to be identified to individual sources. The researchers then compared patterns from those sites with patterns in 170 beams of 6 tree species from 7 great-house structures.

It turns out that the patterns of most beams matched those of trees from the Chuska Mountains (42%) and the Zuni Mountains (28%). Sourcing patterns differed somewhat between tree species. For spruce and fir, the species used in the first strontium-based studies<sup>5,6</sup>, the source deduced by Guiterman *et al.* agreed with that previously identified. Interestingly, the main source shifted with time, from the Zuni Mountains before AD 1020 to the Chuska Mountains thereafter. The strontium studies had not sampled modern trees in the Zuni Mountains because of their greater distance from Chaco, but they agree with the tree-ring-sourcing results that the closer Chuska Mountains were the main source of ponderosa pine, the species used in most Chaco beams. No Chaco beams match the ring patterns of isolated stands of ponderosa pine nearby at Chaco's elevation.

Although Guiterman and colleagues' study solves one mystery about Chaco, it brings others into focus. The first concerns the distance of the wood sources. Archaeologically identified roads radiating from Chaco were presumably the transport routes, but how did people without draught animals transport 5-m-long, 300-kg beams down from mountain forests 1,000 m higher than Chaco, and then 75 km across land, while leaving almost no scratches on the wood?

A second quandary stems from our

knowledge of other ancient construction processes. To build the enormous tall dome of the medieval cathedral in Florence, Italy, heavy weights were raised using a machine consisting of winches, pulleys, gears and wheels, and turned by oxen<sup>9</sup>. The Ancestral Puebloans lacked all five of these, so how did they raise long, heavy roof beams to heights of six storeys?

Third, why did the wood source switch from the Zuni Mountains to the Chuska Mountains a century before Chaco was abandoned by its inhabitants? The switch coincided with a building boom at Chaco, when seven new great houses were erected, and pottery and stone tools also began to be imported from the Chuska Mountains. Might the shift have been caused by deforestation of the Zuni Mountains, by the closer location of the Chuska Mountains to Chaco or by social developments in the source areas?

The seasonal climate, dry conditions and good preservation of archaeological wood in much of western North America make the region well suited for tree-ring-sourcing studies. Now that Guiterman *et al.* have demonstrated the method's value at Chaco, we may hope for many more sourcing studies by archaeologists and historians. ■

**Jared Diamond** is in the Department of Geography, University of California, Los Angeles, Los Angeles, California 90095-1524, USA.  
e-mail: [jdiamond@geog.ucla.edu](mailto:jdiamond@geog.ucla.edu)

1. Lekson, S. H. (ed.) *The Archaeology of Chaco Canyon* (School for Advanced Research Press, 2006).
2. Plog, S. in *Chaco Revisited: New Research on the Prehistory of Chaco Canyon, New Mexico* (eds Heitman, C. C. & Plog, S.) 3–29 (Univ. Arizona Press, 2015).
3. Guiterman, C. H., Swetnam, T. W. & Dean, J. S. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1514272112> (2015).

4. Betancourt, J., Dean, J. & Hull, H. *Am. Antiq.* **51**, 370–375 (1986).  
 5. English, N. B., Betancourt, J. L., Dean, J. S. & Quade, J. *Proc. Natl Acad. Sci. USA* **98**, 11891–11896 (2001).

6. Reynolds, A. et al. *J. Archaeol. Sci.* **32**, 1061–1075 (2005).  
 7. Wills, W. H., Drake, B. L. & Dorschow, W. B. *Proc. Natl Acad. Sci. USA* **111**, 11584–11591 (2014).

8. Bridge, M. J. *Archaeol. Sci.* **39**, 2828–2834 (2012).  
 9. King, R. *La Cupola di Brunelleschi* (BURR Saggi, 2015).

This article was published online on 23 December 2015.

## COSMOLOGY

# Rare isotopic insight into the Universe

**Light isotopes of hydrogen and helium formed minutes after the Big Bang. The study of one of these primordial isotopes, helium-3, has now been proposed as a useful strategy for constraining the physics of the standard cosmological model.**

NIKOS PRANTZOS

The accepted theory of cosmology, known as the standard cosmological model, invokes the existence of a hot early Universe about 13.7 billion years ago. At that time, matter (elementary particles) and radiation (photons) coexisted as an essentially amorphous plasma from which nuclei, atoms, stars and galaxies progressively formed. The observation of ‘relics’ from that period, and their comparison with theoretical predictions, allowed the standard model to be established, and helps scientists to probe the physics of the Universe and to determine the values of its fundamental properties. Writing in *The Astrophysical Journal Letters*, Cooke<sup>1</sup> suggests that observations of the abundance of one such relic, the rare helium isotope <sup>3</sup>He, might provide information about the number of low-mass particle species in the Universe, thus constraining the standard model of nuclear and particle physics.

The hot early Universe left two types of major relic: the faint glow of microwave photons known as the cosmic microwave background (CMB), which is almost the same in all directions of the sky; and the light elements hydrogen and helium. These elements consist of the abundant isotopes <sup>1</sup>H and <sup>4</sup>He, and the rare ones, <sup>2</sup>H (deuterium, also abbreviated to D) and <sup>3</sup>He. All of these isotopes were produced by a process called Big Bang nucleosynthesis (BBN), through nuclear reactions between protons and neutrons during the first few minutes of the hot early Universe.

According to theory, the presently observed properties of the

cosmological relics depend on the physics of the early Universe. For instance, the abundances of primordial deuterium and <sup>3</sup>He depend sensitively on the density of normal (baryonic) matter at that time: the higher the density, the less deuterium and <sup>3</sup>He are produced, because they are more frequently destroyed by primordial nuclear reactions. Similarly, the morphology of the ripples detected in the CMB depends strongly on the cosmic baryonic density.

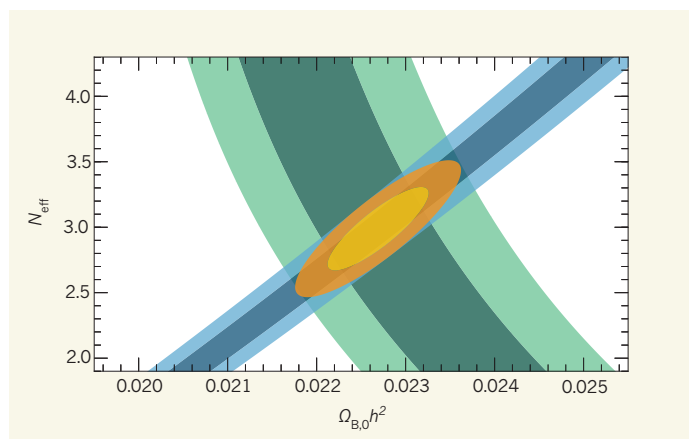
The primordial abundance of <sup>4</sup>He is more sensitive to the expansion rate of the early Universe than to its baryonic density. That

rate depends, in turn, on the number density of photons and other relativistic particles, including electrons, positrons and three flavours of neutrino in the standard model of particle physics. The sum total of all those species is usually parameterized by an ‘effective number of neutrino species’,  $N_{\text{eff}}$ . In the standard model,  $N_{\text{eff}}$  is 3.046, but its value can be different in non-standard models that predict the formation of new particle species.

Theoretical predictions of BBN have improved considerably over the years, and all of the relevant nuclear reaction rates have been measured in the laboratory<sup>2</sup>. But comparison of these predictions with observations requires the primordial abundances of the light nuclei to be reliably established, which is difficult to do. After more than 13 billion years of cosmic evolution, the abundances of all elements in the Universe have been altered by the workings of stars: those of <sup>1</sup>H and deuterium are reduced compared with primordial abundances, because stars ‘burn’ these isotopes in nuclear reactions, whereas the abundances of all other isotopes have steadily increased because they are produced by stars. Regions of the Universe that have evolved very little must therefore be sought if primordial abundances are to be established.

In the case of deuterium, which is the most sensitive chemical probe of baryonic density, observations are made in remote gas clouds more than 10 billion light years (about 3 billion parsecs) away, and therefore more than 10 billion years old. The low content of ‘metals’ (defined by astronomers as elements heavier than helium) in such clouds ensures that their composition is barely affected by stellar activity. The observed isotopic ratio of deuterium to <sup>1</sup>H shows little variation around the average observed value<sup>3</sup>, and points to a baryonic density of 4.5% of the critical cosmic density (the density value that determines whether the Universe is open — expanding forever — or closed). This is in excellent agreement with the value determined from the latest CMB observations by the European Space Agency’s Planck mission<sup>4</sup>.

<sup>4</sup>He is conventionally used as a probe of  $N_{\text{eff}}$ . The abundance of this isotope is measured through the intensity of its emission lines in the gas spectra of nearby



**Figure 1 | Constraining the parameters of the standard cosmological model.** The abundances of nuclei produced during Big Bang nucleosynthesis essentially depend on two parameters: the density of normal (baryonic) matter,  $\Omega_{B,0}$ , and the effective number of neutrino species,  $N_{\text{eff}}$ . The values of  $\Omega_{B,0}$  and  $N_{\text{eff}}$  can be constrained from measurements of the abundance ratio of deuterium to hydrogen (D:<sup>1</sup>H) in near-primordial environments (blue regions indicate constrained values obtained from D:<sup>1</sup>H ratios). Cooke<sup>1</sup> proposes a different method for constraining these parameters, using measurements of the ratio of the yet-to-be-determined primordial abundances of helium-3 and helium-4 isotopes (<sup>3</sup>He:<sup>4</sup>He; green regions indicate constraints based on measurements of <sup>3</sup>He:<sup>4</sup>He values for meteorites that formed at the same time as the Solar System, 4.6 billion years ago). Taken together, the two approaches constrain  $\Omega_{B,0}$  and  $N_{\text{eff}}$  much more than can either individual approach (orange regions indicate combined constraints). Dark and light shades of the coloured regions indicate confidence limits of 68% and 95%, respectively.  $\Omega_{B,0}$  is conventionally expressed as its product with  $h^2$ , where  $h$  is the Hubble parameter divided by 100. (Figure adapted from ref. 1.)





## 50 Years Ago

In the market of Fort Lamy (Chad) one can purchase a greenish edible substance called *Dihe* which is sold as a flat cake. ... It appears to be an alga collected on the bottoms of seasonally dried-up ponds and shallow waters in the north of Lake Chad and consumed by the local population.

However, on arriving in ... Ounianga Kebir ... more than 750 miles to the north-east of Fort Lamy, the botanist was struck by the abundance of a microscopic alga in some lakes. ... Although the local population appears to be unaware that it might have a food value, the botanist ... prepared some cakes according to the recipe obtained.

Both cakes ... are almost exclusively composed of a Cyanophyceae: *Spirulina platensis*. According to chemical analysis it appears that it is a food-plant very rich in proteins.

From *Nature* 8 January 1966

## 100 Years Ago

**The popularisation of Science.**

It is scarcely surprising that scientific knowledge is so little disseminated in this country considering the difficulties which hinder its acquisition. If science is to become widespread, it seems to me essential that it should be democratic both in its higher and in its lower branches. In England, however, science may be said to be aristocratic. Scientific societies demand more or less high subscriptions. Public lectures on science are rarely free. In London an institution exists where advanced lectures are given, but the subscription to which is considerable, and to become members of which people actually have to be recommended—recommended to be allowed to learn!

From *Nature* 6 January 1916

galaxies that have low metal content, but those measurements are affected by systematic uncertainties. Even worse, the latest analyses point to a primordial  $^4\text{He}$  abundance that seems to be significantly higher than the one suggested by the Planck mission's CMB study<sup>5,6</sup>.

To resolve this problem and to reduce the uncertainties, Cooke proposes that the usually neglected primordial isotope  $^3\text{He}$  should be included in the analyses. According to BBN theory, the ratio of the primordial abundance of  $^3\text{He}$  to that of  $^4\text{He}$  depends on both  $N_{\text{eff}}$  and the cosmic baryonic density, in a way that is opposite to the dependence of the ratio of deuterium to  $^1\text{H}$ ; that is,  $^3\text{He}:^4\text{He}$  decreases with  $N_{\text{eff}}$ , whereas  $\text{D}:^1\text{H}$  increases. So, by combining analyses of both the hydrogen and helium isotopic ratios, the value of  $N_{\text{eff}}$  can be constrained better than by using either the abundance of  $^4\text{He}$  or the  $\text{D}:^1\text{H}$  ratio alone (Fig. 1).

Implementing this idea is far from trivial, however, on both observational and theoretical grounds. First, uncertainties in nuclear-reaction rates will have to be further reduced to make  $^3\text{He}$  a useful probe for precision cosmology. Second, unlike deuterium, which is always destroyed by stars,  $^3\text{He}$  is produced by low-mass stars but destroyed by higher-mass ones, to a poorly known extent. This makes it difficult to determine its primordial abundance unambiguously, even by looking in low-metallicity environments.

Moreover,  $^3\text{He}$  is 10,000 times less abundant than  $^4\text{He}$ , and so its weak emission line

will be hard to identify in the background of the much brighter  $^4\text{He}$  line — especially if the latter is broadened by rapid thermal or turbulent motions of the emitting gas. A statistically significant detection of  $^3\text{He}$  would require a high signal-to-noise ratio, of more than 500. This will be obtainable only using the next generation of telescopes, which will have mirrors 30 metres or more in diameter.

Nevertheless, Cooke's suggestion is of great interest, because the standard cosmological model should be checked as accurately as possible with every available method, in view of its prominent role in modern physics. In particular, Cooke's strategy should allow potential departures from the standard model to be probed in a complementary way to existing strategies. ■

**Nikos Prantzos** is at the *Institut d'Astrophysique de Paris, 75014 Paris, France, and the Université Pierre et Marie Curie, Paris.*  
e-mail: prantzos@iap.fr

1. Cooke, R. *Astrophys. J.* **812**, L12 (2015).
2. Cyburt, R. H., Fields, B. D., Olive, K. A. & Yeh, T.-H. Preprint at <http://arxiv.org/abs/1505.01076> (2015).
3. Cooke, R. J., Pettini, M., Jorgenson, R. A., Murphy, M. T. & Steidel, C. C. *Astrophys. J.* **781**, 31 (2014).
4. Planck Collaboration. Preprint at <http://arxiv.org/abs/1502.01589> (2015).
5. Aver, E., Olive, K. A. & Skillman, E. D. *J. Cosmol. Astropart. Phys.* **7**, 011 (2015).
6. Izotov, Y. I., Thuan, T. X. & Guseva, N. G. *Mon. Not. R. Astron. Soc.* **445**, 778 (2014).

This article was published online on 9 December 2015.

### CANCER

## Oncogene brought into the loop

**Analysis of the 3D structure of DNA in tumour cells reveals how mutations in the IDH1 gene, and associated changes in methyl groups attached to DNA, elevate the expression of cancer-promoting genes. SEE LETTER P.110**

**MATTHEW R. GRIMMER & JOSEPH F. COSTELLO**

The discovery in the late 2000s that mutations in the gene that encodes the enzyme isocitrate dehydrogenase 1 (*IDH1*) are often associated with glioma, the most common form of brain cancer, was unexpected and tantalizing<sup>1,2</sup>. The *IDH1* protein is involved in the citric-acid cycle — a metabolic process that is used by nearly all cells to generate energy, and that in 2008 had only recently been connected to cancer<sup>3,4</sup>. The discovery therefore supported the long-standing theory that altered metabolism could transform normal cells into cancerous

ones. On page 110 of this issue, Flavahan *et al.*<sup>5</sup> report that an abnormal metabolite generated by mutant *IDH1* may drive cancer primarily by altering the 3D conformation of DNA.

Mutant *IDH1* converts the citric-acid-cycle molecule isocitrate into an abnormal metabolite that inhibits TET enzymes<sup>6</sup>, which remove methyl groups from DNA. The presence of methyl groups can alter gene expression by preventing some proteins from binding DNA, and an excess of methyl groups in promoter sequences (which drive gene expression) can silence tumour-suppressor genes, leading to cancer. It has been suggested<sup>7</sup> that inhibition of TET enzymes

leads to such hypermethylation in *IDH1*-mutant tumours. However, promoter hypermethylation in these tumours is not generally correlated with changes in gene expression<sup>8</sup>, suggesting that cancer-associated changes in methylation may occur at other DNA sequences.

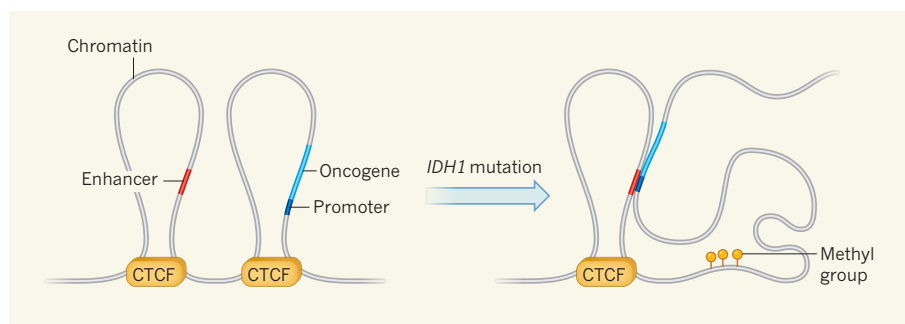
In addition to promoter regions, gene expression can be regulated by the 3D structure of chromatin (the complex in which DNA is wound around histone proteins for packaging in the cell). Chromatin structure is exceptionally intricate, and is defined in part by evolutionarily conserved loops called topologically associated domains (TADs). Interactions between DNA sequences — for instance, those that bring promoters into contact with distant enhancer elements to activate gene expression — are more common within than between TADs, and there is evidence<sup>9</sup> that gene expression is coordinated in these loops.

TADs are insulated from one another by DNA-binding proteins such as the CTCF-binding factor (CTCF). Deletion of the DNA sequence encompassing one CTCF binding site has been shown to cause changes in TAD structure and gene expression that lead to limb malformations<sup>10</sup>, highlighting the importance of maintaining these boundaries. Notably, CTCF binding is sensitive to changes in DNA methylation<sup>11,12</sup>.

Flavahan *et al.* demonstrated that a subset of CTCF binding sites is methylated in *IDH1*-mutant gliomas, and that CTCF binding at these sites is subsequently reduced. Leveraging gene-expression data from hundreds of gliomas and normal brain specimens, and using 3D chromosome-conformation data from various cell lines, the authors found previously unknown gene-expression correlations between TADs in *IDH1*-mutant gliomas, suggesting that TAD borders are disrupted.

Hundreds of the pairs of genes that are correlated in the mutant cells straddle a disrupted TAD border. Of these, *PDGFRA* and *FIP1L1* are among the most highly expressed. *PDGFRA* is an appealing candidate for further study, because it is a well-documented oncogene (it promotes cancer when mutationally activated or overexpressed) and is amplified genetically in some 20% of advanced (high-grade) gliomas<sup>13</sup>. The authors find that, in *IDH1*-mutant gliomas, which are low grade, the CTCF site at the TAD boundary between *PDGFRA* and *FIP1L1* is methylated and CTCF binding is reduced. Thus, an increase in *PDGFRA* expression, although arising through different mechanisms in low- and high-grade tumours, may be a common theme in glioma.

Flavahan and colleagues showed that, in glioma cells in which *IDH1* is not mutated, the *PDGFRA* promoter strongly interacts with its own enhancer. The interaction patterns are



**Figure 1 | Breaking down boundaries to cancer.** Structural boundaries between regions of chromatin (the complex of DNA and proteins in which DNA is packaged in the nucleus) define loops called topologically associated domains (TADs), within which gene activity is coordinated. DNA binding by the insulator protein CTCF separates these domains. Flavahan *et al.*<sup>5</sup> provide evidence that CTCF insulation prevents the activation of oncogenes (genes whose hyperactivity promotes cancer) by distant enhancer elements from different TADs. The authors find that mutations in the gene *IDH1* increase the number of methyl groups that are attached to CTCF binding sites, reducing CTCF–DNA binding. This breaks down the TAD border structure, allowing aberrant association between enhancers and the promoter regions of oncogenes. Oncogene expression is subsequently amplified, leading to cancer.

markedly different in *IDH1*-mutant tumours. Here, there is a strong interaction between the *PDGFRA* promoter and the unrelated enhancer of *FIP1L1*, despite the fact that these two genetic elements are separated by almost 900,000 base pairs. This aberrant interaction is approximately five times stronger than that between the *PDGFRA* promoter and its own enhancer. Together, these results suggest that disruption of a boundary element by hypermethylation allows a potent *FIP1L1* enhancer to interact with the *PDGFRA* promoter, increasing gene expression (Fig. 1).

To confirm that DNA hypermethylation is responsible for the elevated *PDGFRA* expression that they observed, the authors treated *IDH1*-mutant cells with a drug that reduces DNA methylation. In agreement with their hypothesis, the treatment reduced methylation of the relevant CTCF binding site, increasing CTCF binding and reducing *PDGFRA* expression. Conversely, experimental disruption of the CTCF binding site in cells that lacked the *IDH1* mutation led to increased *PDGFRA* expression. The altered expression presumably occurs because of changes in enhancer–promoter interactions, but this was not tested directly. Elevated *PDGFRA* expression doubled cell growth compared with untreated cells. This suggests that the increased *PDGFRA* protein in *IDH1*-mutant glioma cells provides a selective growth advantage over cells lacking the mutation.

Flavahan and colleagues' study focuses on one CTCF site out of hundreds, so other oncogenes might also be activated by newly formed enhancer–promoter interactions in *IDH1*-mutant tumours. Many newly activated genes may also be 'passenger' events, which have no functional consequences. The methylation states of CTCF sites and the activity of enhancers vary widely across cell types, suggesting that 3D chromosome-conformation analysis of high-grade gliomas, colorectal cancers, lymphomas, leukaemias and other *IDH1*-mutant

cancers could reveal different targets of genomic hypermethylation. These targets may also include those normally bound by methylation-sensitive factors other than CTCF.

Consistent with the fact that DNA methylation is highly stable, aberrant hypermethylation persists in *IDH1*-mutant tumours after treatment with an inhibitor of mutant *IDH1* (ref. 14). Assuming that hypermethylation is involved in the transition to cancer, as is strongly suggested by the current study, such stability could pose a challenge for the success of *IDH1*-inhibitor treatments in patients. Unravelling the effects of DNA hypermethylation on gene dysregulation will lead to a more complete survey of the forces downstream of TET and other enzymes that drive the evolution of *IDH1*-mutant cancer cells<sup>15</sup>. Flavahan and colleagues' study provides a fresh perspective on which to base such future analyses. ■

**Matthew R. Grimmer and Joseph F. Costello** are in the Department of Neurological Surgery, University of California, San Francisco, San Francisco, California 94158, USA.  
e-mails: joseph.costello@ucsf.edu; matthew.grimmer@ucsf.edu

1. Parsons, D. W. *et al.* *Science* **321**, 1807–1812 (2008).
2. Yan, H. *et al.* *N. Engl. J. Med.* **360**, 765–773 (2009).
3. Selak, M. A. *et al.* *Cancer Cell* **7**, 77–85 (2005).
4. The Multiple Leiomyoma Consortium. *Nature Genet.* **30**, 406–410 (2002).
5. Flavahan, W. A. *et al.* *Nature* **529**, 110–114 (2015).
6. Xu, W. *et al.* *Cancer Cell* **19**, 17–30 (2011).
7. Turcan, S. *et al.* *Nature* **483**, 479–483 (2012).
8. Noshmeh, H. *et al.* *Cancer Cell* **17**, 510–522 (2010).
9. Nora, E. P. *et al.* *Nature* **485**, 381–385 (2012).
10. Lupiáñez, D. G. *et al.* *Cell* **161**, 1012–1025 (2015).
11. Bell, A. C. & Felsenfeld, G. *Nature* **405**, 482–485 (2000).
12. Hark, A. T. *et al.* *Nature* **405**, 486–489 (2000).
13. Phillips, J. J. *et al.* *Brain Pathol.* **23**, 565–573 (2013).
14. Rohle, D. *et al.* *Science* **340**, 626–630 (2013).
15. Koivunen, P. *et al.* *Nature* **483**, 484–488 (2012).

**J.F.C. declares competing financial interests.** See online article for details.

This article was published online on 23 December 2015.



# Autophagy maintains stemness by preventing senescence

Laura García-Prat<sup>1</sup>, Marta Martínez-Vicente<sup>2\*</sup>, Eusebio Perdiguero<sup>1</sup>, Laura Ortet<sup>1</sup>, Javier Rodríguez-Ubrea<sup>3</sup>, Elena Rebollo<sup>4</sup>, Vanessa Ruiz-Bonilla<sup>1</sup>, Susana Gutarra<sup>1</sup>, Esteban Ballestar<sup>3</sup>, Antonio L. Serrano<sup>1</sup>, Marco Sandri<sup>5,6\*</sup> & Pura Muñoz-Cánoves<sup>1,7</sup>

**During ageing, muscle stem-cell regenerative function declines. At advanced geriatric age, this decline is maximal owing to transition from a normal quiescence into an irreversible senescence state. How satellite cells maintain quiescence and avoid senescence until advanced age remains unknown. Here we report that basal autophagy is essential to maintain the stem-cell quiescent state in mice. Failure of autophagy in physiologically aged satellite cells or genetic impairment of autophagy in young cells causes entry into senescence by loss of proteostasis, increased mitochondrial dysfunction and oxidative stress, resulting in a decline in the function and number of satellite cells. Re-establishment of autophagy reverses senescence and restores regenerative functions in geriatric satellite cells. As autophagy also declines in human geriatric satellite cells, our findings reveal autophagy to be a decisive stem-cell-fate regulator, with implications for fostering muscle regeneration in sarcopenia.**

The regenerative capacity of skeletal muscle relies on long-lived Pax7-expressing muscle stem cells (called satellite cells), which are normally in quiescence (a G0 reversible arrest state). In response to tissue damage, these cells activate, enter the cell cycle and either expand and form new myofibres or self-renew to restore the quiescent satellite cell pool<sup>1–4</sup>. Quiescence therefore appears to be a simple way of functionally maintaining the stem-cell population throughout life in the absence of regenerative demand, particularly in tissues with little turnover, such as skeletal muscle.

Sarcopenia, the age-related loss of skeletal muscle mass and function, is maximal at geriatric age. At this last stage of life, skeletal muscle shows a profound regenerative impairment that contributes to the individual's physical incapacitation. Both changes in the environment (such as inflammatory status) and/or satellite-cell-intrinsic mechanisms associated to ageing may contribute to this regenerative decline<sup>5,6</sup>. Recent studies have demonstrated that aged skeletal muscles fail to retain stem-cell quiescence<sup>7–9</sup>. Both the number and the functionality of muscle stem cells decline with ageing<sup>7–13</sup>, with satellite cells switching from a quiescence to a pre-senescence state in sarcopenic muscle at geriatric age<sup>8</sup>. How satellite cells maintain quiescence during their long life and avoid acquisition of the senescence program until advanced age is largely unknown.

Using physiologically aged mice, we show that quiescent muscle stem cells preserve their integrity over time through active maintenance of organelle and protein homeostasis (proteostasis) as a cellular quality control mechanism. We demonstrate that these dormant stem cells display continuous basal macroautophagy (hereafter referred to as 'autophagy'; that is, the process for degradation of long-lived proteins and damaged organelles in lysosomes<sup>14,15</sup>). This activity declines during ageing. Physiological decline of autophagy in old satellite cells or its genetic impairment in young cells, results in toxic cellular waste accumulation, resulting in entry into senescence.

Our studies indicate that muscle stem cells preserve their G0-reversible quiescence state from entering a G0-irreversible

senescence state through autophagy. Genetic and pharmacological regimes that reinstall basal autophagy in geriatric mice reversed stem-cell senescence and restored regeneration, which has implications for the use of regenerative medicine in sarcopenia.

## Impaired autophagy in aged satellite cells

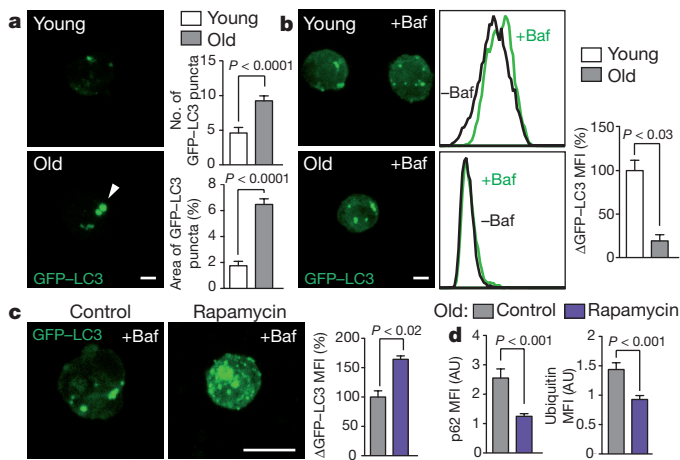
We interrogated the transcriptomes of quiescent satellite cells compared to activated cells for changes in proteostasis genes<sup>16–18</sup> and uncovered autophagy as the most prevalent pathway in the quiescent state (Extended Data Fig. 1a and Supplementary Table 1). *K*-means clustering analysis revealed an age-associated downregulation of autophagic genes in quiescence (Extended Data Fig. 1b and Supplementary Table 1).

Autophagy is an evolutionary conserved process of self-degradation of cellular components (organelles, cytosol portions and misfolded proteins) by autophagosomes, which are delivered to the lysosomal machinery, thus preventing waste accumulation<sup>14,15</sup>, and this process has been implicated in ageing of different model organisms<sup>14,15,19,20</sup>. To investigate the occurrence of autophagy in quiescent muscle stem cells we used green fluorescent protein (GFP)–LC3 (a well-known marker of autophagosomes) transgenic mice<sup>21,22</sup>. Quiescent satellite cells were isolated by fluorescence-activated cell sorting (FACS) (Extended Data Fig. 1c) from the resting muscle of young (3 months) and old (20–24 months) GFP–LC3 mice. Punctate GFP–LC3 signal was found in young cells, and this was increased in old cells (Fig. 1a, Extended Data Fig. 1d and Supplementary Videos 1 and 2). We next used the autophagy-flux inhibitor bafilomycin, which prevents lysosome degradation, thus increasing punctate GFP–LC3 exclusively when autophagy is active<sup>23</sup>. Bafilomycin treatment demonstrated that—in contrast to the result for young cells—old satellite cells lacked the capacity for further autophagosome formation, as monitored by GFP–LC3 fluorescence levels (Fig. 1b). These results indicate constitutive autophagic activity in young quiescent satellite cells and impaired autophagic activity during ageing. Fluorescence, transmission-electron microscopy and western

<sup>1</sup>Cell Biology Group, Department of Experimental and Health Sciences, Pompeu Fabra University (UPF), CIBER on Neurodegenerative diseases (CIBERNED), E-08003 Barcelona, Spain.

<sup>2</sup>Neurodegenerative Diseases Research Group, Vall d'Hebron Research Institute-CIBERNED, E-08035 Barcelona, Spain. <sup>3</sup>Chromatin and Disease Group, Cancer Epigenetics and Biology Programme (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, E-08907 Barcelona, Spain. <sup>4</sup>Advanced Fluorescence Microscopy Unit, Molecular Biology Institute of Barcelona (IBMB-CSIC), E-08028 Barcelona, Spain. <sup>5</sup>Department of Biomedical Science, University of Padova, 35100 Padova, Italy. <sup>6</sup>Telethon Institute of Genetics and Medicine (TIGEM), 80131 Napoli, Italy. <sup>7</sup>ICREA, E-08908 Barcelona, Spain.

\*These authors contributed equally to this work.



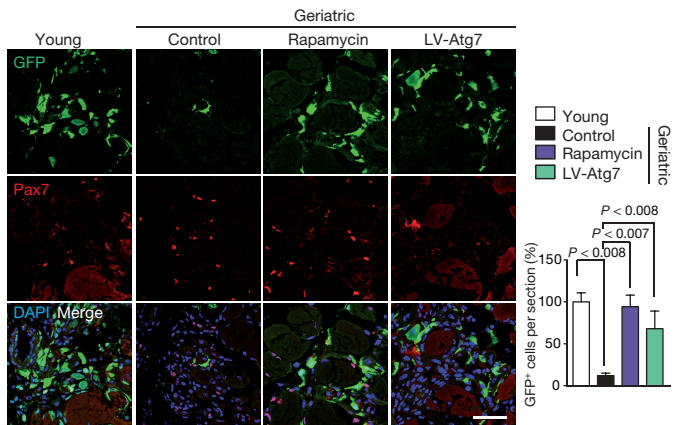
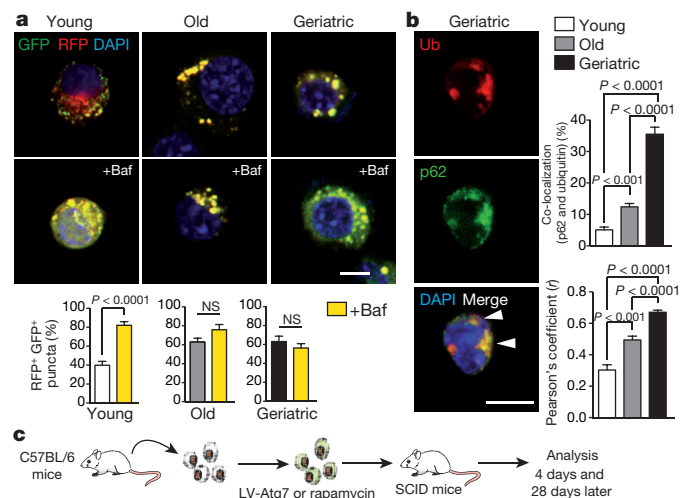
**Figure 1 | Altered basal autophagy in ageing muscle stem cells.**

**a**, Number and area of punctate GFP-LC3 in quiescent satellite cells. Arrowheads, autophagic vesicles. **b**, Autophagy flux in cells from **a**. Cells were treated with vehicle or bafilomycin (+Baf) for 4 h before analysis. The change in mean fluorescence intensity (MFI) of GFP-LC3 in cells treated with bafilomycin is shown. **c**, Autophagy flux in quiescent satellite cells from old GFP-LC3 mice treated for two weeks with rapamycin or vehicle control. Satellite cells  $\pm$  bafilomycin treatment as in **b**. **d**, p62 and ubiquitin (Ub) MFI from cells of the GFP-LC3 mice treated as in **c**. AU, arbitrary units. Data show mean  $\pm$  s.e.m. Comparisons by two-sided Mann-Whitney *U*-test. Sample numbers were  $n = 51$  (young),  $n = 106$  (old) cells analysed from 3 animals for **a**;  $n = 60,000$  cells from 3 animals for **b**;  $n = 60,000$  cells from 3 animals for **c**;  $n = 36$  (control), 39 (rapamycin treated) cells from 3 animals for **d**. The *z* projections of representative images are shown. Scale bars are all 1.5  $\mu$ m, apart from 5  $\mu$ m for **c**.

blotting analyses indicated common traits of deficient autophagy in old satellite cells, including the accumulation of autophagic vesicles (Extended Data Fig. 1e, f), aggregates forming of p62 (a protein regulating autophagic clearance of dysfunctional organelles or aggregates), ubiquitin (Ub)-positive inclusions (Extended Data Fig. 1g), reduced LC3II accumulation after bafilomycin treatment (Extended Data Fig. 1h). A two-week-treatment in old mice with rapamycin (or spermidine), well-known autophagy-inducing regimes<sup>24,25</sup>, restored basal autophagy in stem cells (Fig. 1c, Extended Data Fig. 1i and Supplementary Videos 3 and 4) and reduced protein and organelle aggregates (Fig. 1d and Extended Data Fig. 1j).

### Restoring autophagy prevents senescence

Satellite cells enter a senescent state when they reach a geriatric age (over 28 months in mice)<sup>8,26</sup>. We investigated whether dysregulated basal autophagy may underlie the loss of bona fide quiescence. Using an mRFP-GFP-LC3 construct<sup>27</sup> (a tandem fluorescent-tagged LC3 reporter containing monomeric red fluorescent protein (mRFP) and GFP), transfected into young, old and geriatric satellite cells, analysing these samples in combination with bafilomycin treatment, we found a higher blockade of autophagic flux in geriatric than old cells, with respect to young cells (the blockade is geriatric > old > young). In the absence of bafilomycin, red LC3 puncta (mature autolysosomes) were only abundant in young cells. Bafilomycin treatment induced yellow LC3 puncta (non-fused autophagosomes) accumulation in young cells, which was blunted in old and geriatric cells (Fig. 2a). Geriatric satellite cells also showed increased co-localization of p62-ubiquitin aggregates in non-degraded autophagosomes (Fig. 2b). As p62 marks damaged organelles for degradation by selective autophagy, whereas ubiquitin marks substrates for their degradation by either the ubiquitin-proteasome system (UPS) or selective autophagy, the increased signal of both proteins and their co-localization demonstrates that the autophagic defect in these cells is due, at least in part, to a block in autophagosomal or lysosomal clearance.



**Figure 2 | Defective autophagy causes numerical and functional satellite cell decline in ageing.**

**a**, The mRFP-GFP-LC3 plasmid was transfected into young (3 months), old (24 months) and geriatric (28 months) satellite cells to enable the detection of autophagosomes (yellow) and their maturation into autolysosomes (red) in the presence or absence of bafilomycin treatment (as in Fig. 1b). The graph indicates the percentage of double-positive puncta (RFP<sup>+</sup>/GFP<sup>+</sup>) (autophagosomes) out of total puncta (RFP<sup>+</sup>/GFP<sup>+</sup>, RFP<sup>+</sup>, autophagosomes and autolysosomes). **b**, Quantification of p62 and ubiquitin aggregates in quiescent satellite cells from **a**. Co-localization staining area with respect to total cellular area. Pearson's coefficient (*r*) indicates the correlation of intensity values of red and green pixels in dual-channel images. Arrowheads indicate co-localization. **c**, An equal number of LV-GFP-infected satellite cells from young or geriatric mice, treated for 48 h with rapamycin or LV-Atg7 infected (or controls), were transplanted into an injured mouse muscle, and analysed 4 or 28 days later (for analysis on day 28, see Extended Data Fig. 2i, j). Analysis on day 4 of GFP and Pax7 immunostaining. Quantification of GFP<sup>+</sup> cells per muscle field. Values relative to transplanted young cells (100%). Data show mean  $\pm$  s.e.m. Comparisons by two-sided Mann-Whitney *U*-test. The sample numbers were  $n = 21$  (young),  $n = 19$  (young, +bafilomycin),  $n = 30$  (old),  $n = 15$  (old, +bafilomycin),  $n = 21$  (geriatric) and  $n = 15$  (geriatric, +bafilomycin) cells analysed from 3 animals for **a**;  $n = 35$  (young),  $n = 66$  (old) and  $n = 104$  (geriatric) cells from 3 animals for **b**;  $n = 5$  engraftments per group for **c**. Representative images are shown. The *z* projections of representative images are shown. Scale bars are all 5  $\mu$ m, apart from 50  $\mu$ m for **c**.

To investigate whether restoring autophagy could rescue the cell-intrinsic irreversible cell cycle and regenerative block of geriatric cells, we engrafted freshly isolated GFP-labelled young and geriatric satellite cells (pre-treated with rapamycin or the control vehicle) into pre-injured muscles of young recipient mice. Autophagy reactivation significantly restored expansion of geriatric cells (expressing Pax7, Ki67, MyoD or myogenin (Mgn, also known as Myog)) after a four-day engraftment (Fig. 2c and Extended Data Fig. 2a–c) and prevented senescence



(geroconversion), as shown by p16<sup>INK4a</sup> and  $\gamma$ H2AX reduction (Extended Data Fig. 2d). Rapamycin (or spermidine) treatment also decreased geriatric senescent cells (senescence-associated  $\beta$ -galactosidase-positive (SA- $\beta$ -gal<sup>+</sup>)) (Extended Data Fig. 2e, f) and re-established proliferation (Extended Data Fig. 2e). A genetic approach to enhance autophagy, by overexpressing Atg7 (crucial for autophagosome formation) (Fig. 2c and Extended Data Fig. 2g, h) rescued the proliferative defect, while reducing senescence (Extended Data Fig. 2e). Furthermore, satellite cell transplantation and whole-muscle graft experiments demonstrated that the introduction of Atg7 alone in geriatric satellite cells rescued their intrinsic regenerative capacity, allowing the formation of new muscle fibres (Fig. 2c and Extended Data Fig. 2i, j).

### Atg7 loss causes stem-cell senescence

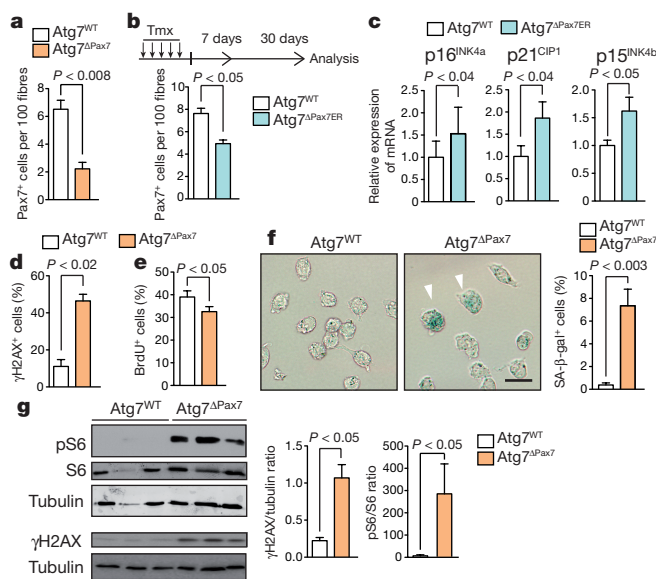
To investigate if basal autophagy disruption causally breaks quiescence, we intercrossed Atg7-floxed mice with Pax7-Cre and Pax7-Cre<sup>ER</sup> mice, to impair autophagy in Pax7-expressing cells either constitutively (Atg7 <sup>$\Delta$ Pax7</sup>) or inducibly (Atg7 <sup>$\Delta$ Pax7ER</sup>) after tamoxifen administration. Intercrossing Atg7 <sup>$\Delta$ Pax7</sup> with GFP-LC3 mice (Atg7 <sup>$\Delta$ Pax7</sup>:GFP-LC3) resulted in the loss of autophagosomes in quiescent Atg7 null satellite cells (Extended Data Fig. 3a, b). The satellite cell pool was severely reduced in Atg7 <sup>$\Delta$ Pax7</sup> mice (Fig. 3a and Extended Data Fig. 3c). Tamoxifen administration to three-month-old Atg7 <sup>$\Delta$ Pax7ER</sup> mice led to satellite cell loss after 30 days (Fig. 3b), indicating that basal autophagy is required for both establishment and maintenance of the adult quiescent stem-cell population. The remaining Atg7 <sup>$\Delta$ Pax7ER</sup> satellite cells showed unexpected signs of premature ageing including induction of p16<sup>INK4a</sup>, p21<sup>CIP1</sup> and p15<sup>INK4b</sup> and DNA damage ( $\gamma$ H2AX<sup>+</sup> cells) (Fig. 3c, d and Extended Data Fig. 3d). Atg7 <sup>$\Delta$ Pax7</sup> satellite cells did not undergo

mitotic or myogenic differentiation pathways (Extended Data Fig. 3e). Thus, loss of autophagy with ageing may be the cause underlying the age-associated numerical decline in muscle stem cells<sup>7,8,10–13</sup>.

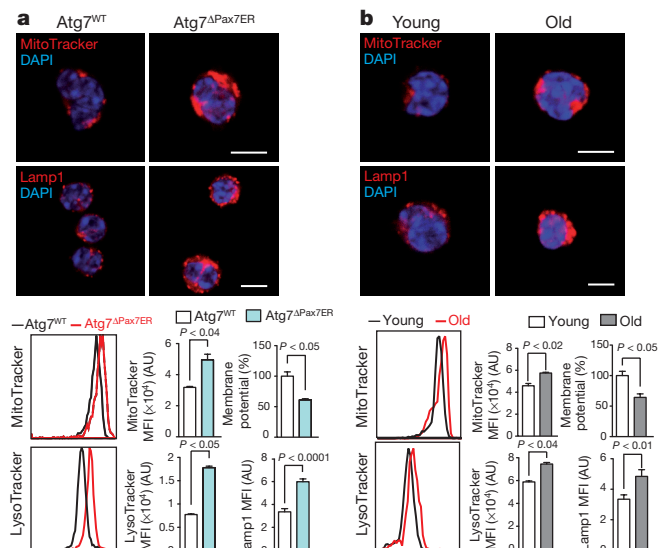
In response to muscle injury, Pax7<sup>+</sup> cells from young Atg7 <sup>$\Delta$ Pax7</sup> mice showed reduced activation and expansion capacity (Fig. 3e and Extended Data Fig. 3f), and accelerated entry into deep senescence<sup>26,28</sup> (geroconversion<sup>29</sup>) *in vivo* and *in vitro*, as demonstrated by SA- $\beta$ -gal-staining (Fig. 3f), and increased expression of  $\gamma$ H2AX and phosphorylated S6, and also evidence of regenerative failure, shown by reduced cell proliferation and decreased size of regenerating fibres (Fig. 3g and Extended Data Fig. 3g–i). Confirming the cell-intrinsic regenerative failure, fewer GFP<sup>+</sup> fibres derived from Atg7 null satellite cells were found in transplantation experiments (Extended Data Fig. 3m, n), and this failure could not be rescued by rapamycin (or spermidine) (Extended Data Fig. 3m–o).

### Altered mitophagy and increased ROS induce senescence

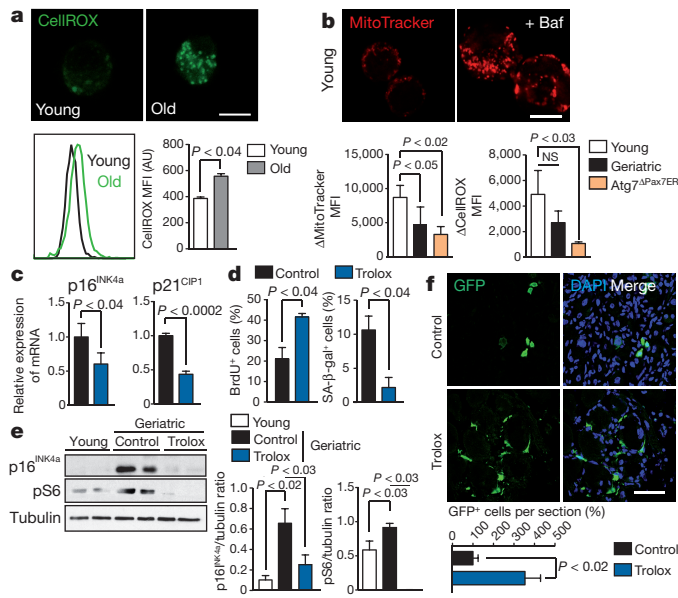
We next investigated how loss of autophagy in young quiescent satellite cells induced premature ageing. Genetic disruption of autophagy in satellite cells resulted in a similar phenotype to that observed in aged cells with rapid accumulation of p62 and ubiquitin-positive aggregates, and also mitochondria and lysosomes (MitoTracker, and LysoTracker and Lamp1) (Fig. 4a, b and Extended Data Fig. 4a). There was also a lower proportion of healthy mitochondria in old (and Atg7 <sup>$\Delta$ Pax7ER</sup>) satellite cells, as revealed by reduced membrane potential (a lower mean fluorescence intensity ratio of the active mitochondria labelling fluorescent dye TMRM to MitoTracker green) (Fig. 4a, b). Furthermore, mitophagy (the cellular capacity to clear damaged mitochondria by autophagy) was defective in geriatric satellite cells, as indicated by mitochondria accumulation inside autophagosomes or lysosomes (through co-localization of mitochondrial TOM20 and lysosomal Lamp-1 markers) (Extended Data Fig. 4b). *In vivo* rapamycin (or spermidine) treatment of geriatric mice restored mitophagy in satellite cells (Extended Data Fig. 4b–f). Consistent with age-impaired mitophagy, young, but not geriatric, cells were capable of eliminating carbonyl



**Figure 3 | Genetic impairment of autophagy disrupts satellite cell homeostasis.** **a**, Satellite cell quantification by analysis of Pax7 immunostaining of muscles in three-month-old Atg7<sup>WT</sup> and Atg7 <sup>$\Delta$ Pax7</sup> mice. **b**, Schematic of the mouse tamoxifen (Tmx) treatment and satellite cell analysis. Satellite cell quantification in Atg7<sup>WT</sup>/Atg7 <sup>$\Delta$ Pax7ER</sup> mice on day 30 after tamoxifen treatment as in **a**. **c**, RT-qPCR of senescence markers (7 days after tamoxifen treatment) in cells from **b**. **d**, The percentage of co-localizing  $\gamma$ H2AX<sup>+</sup> cells out of the total Pax7<sup>+</sup> cells in **a**. **e**, **f**, Quantification of BrdU<sup>+</sup> (**e**) and SA- $\beta$ -gal<sup>+</sup> (**f**) cells isolated from mice in **a**. The arrowhead indicates positive staining. **g**, Western blotting of cells from **a**; for a full scan of the gel see Supplementary Fig. 1. pS6, phosphorylated S6. Data show mean  $\pm$  s.e.m. Comparisons by two-sided Mann–Whitney *U*-test. Sample numbers were: *n* = 5 animals per group for **a**; *n* = 3 animals per group for **b–g**. Representative images are shown. The *z* projections of representative images are shown. Scale bars, 250  $\mu$ m.



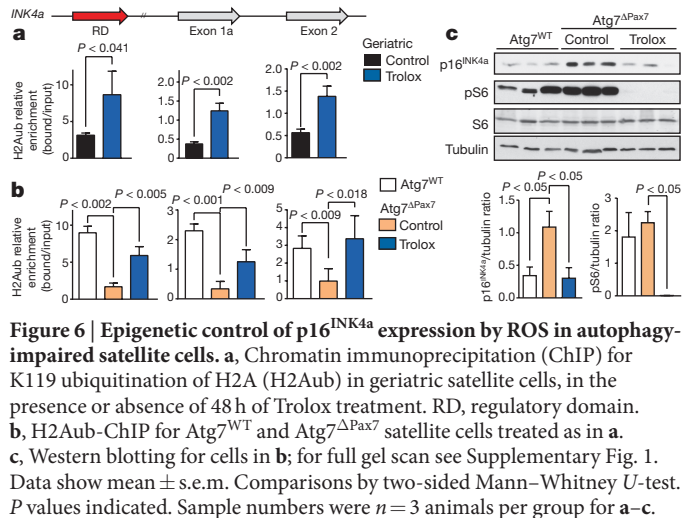
**Figure 4 | Autophagy loss results in mitochondrial dysfunction and accumulation of organelles, proteins and ROS.** **a**, Lysosomal (Lamp1 and LysoTracker) and mitochondrial (MitoTracker) quantification in satellite cells from Atg7<sup>WT</sup> and Atg7 <sup>$\Delta$ Pax7ER</sup> mice, one month after tamoxifen treatment. Membrane potential monitored as the mean fluorescent intensity ratio of TMRM to MitoTracker Green. **b**, Similar quantification as in **a** for young and old satellite cells. Membrane potential analysis as in **a**. Data show mean  $\pm$  s.e.m. Comparisons by two-sided Mann–Whitney *U*-test. The sample numbers were *n* = 60,000 cells analysed from 3 animals for **a** and **b**. Representative images are shown. The *z* projections of representative images are shown. Scale bars, 5  $\mu$ m.



**Figure 5 | ROS inhibition prevents senescence in aged satellite cells.** **a**, ROS quantification in young and old satellite cells by CellROX flow cytometry. **b**, Mitochondria (MitoTracker) and ROS (CellROX) in satellite cells, ± 24 h bafilomycin treatment. Results represent increased MFI in the presence of bafilomycin. Representative images of young satellite cells. **c**, RT-qPCR of senescence markers ± Trolox. **d**, Quantification of BrdU<sup>+</sup> and SA-β-gal<sup>+</sup> cells from **c**, pretreated with Trolox (or controls that were not treated with Trolox), and cultured for 96 h. **e**, Western blotting in cells ± Trolox; for full-scan of gel see Supplementary Fig. 1. **f**, Geriatric cells ± Trolox 48 h pre-treatment, were transplanted and analysed as in Fig. 2c. Data show mean ± s.e.m. Comparisons by two-sided Mann-Whitney *U*-tests. Sample numbers were *n* = 60,000 cells analysed from 3 animals for **a** and **b**; *n* = 3 animals per group for **c**–**e**; *n* = 4 engraftments per group for **f**. Representative images are shown. The *z* projections of representative images are shown. Scale bars are all 5 μm apart from 50 μm for **f**.

cyanide 3-chlorophenylhydrazine (CCCP)-damaged mitochondria (Extended Data Fig. 4d, e).

Next, we analyzed how altered mitophagy led to satellite cell senescence with ageing. We detected higher levels of reactive-oxygen species (ROS), parkin (marking damaged mitochondria for degradation by mitophagy), and DNA-damage markers in Atg7-deficient satellite cells (Fig. 3d, g and Extended Data Figs 3h and 5a, b), associated with p16<sup>INK4a</sup> and pS6 induction (Fig. 3g and Extended Data Figs 3g and 4h). Higher ROS labelling and ROS-mitochondria co-localization were also observed in geriatric satellite cells, correlating with impaired mitophagic flux (Fig. 5a and Extended Data Fig. 4g). Bafilomycin-induced autophagy block caused greater mitochondrial accumulation in young cells, compared with geriatric and Atg7<sup>ΔPax7ER</sup> cells, paralleling the ROS increase (Fig. 5b). To address the role of ROS in impaired autophagy, we inhibited it with Trolox (a vitamin E analogue) (Extended Data Fig. 5c). Trolox treatment of old GFP-LC3 mice increased GFP-LC3 puncta (after bafilomycin treatment) and reduced p62 and ubiquitin aggregates and mitochondria-ROS co-localization in GFP-LC3 satellite cells (Extended Data Figs 4g and 5d). Attenuation of autophagic block by ROS inhibition was further confirmed in bafilomycin-treated aged cells through LC3-II accumulation (Extended Data Fig. 5e, f) and an mRFP-GFP-LC3 tandem reporter, which detected reduced autophagosomes (RFP<sup>+</sup>/GFP<sup>+</sup> puncta) and rescued autophagic flux (Extended Data Fig. 5g). Trolox treatment prevented the appearance of senescence markers (Fig. 5c–e), restored the expansion (Fig. 5d), and rescued the cell-intrinsic proliferative and regenerative defect of geriatric satellite cells after transplantation (Fig. 5f and Extended Data Fig. 5h). Thus, increased ROS, resulting from impaired autophagy, drive satellite cell senescence in aged cells.



**Figure 6 | Epigenetic control of p16<sup>INK4a</sup> expression by ROS in autophagy-impaired satellite cells.** **a**, Chromatin immunoprecipitation (ChIP) for K119 ubiquitination of H2A (H2Aub) in geriatric satellite cells, in the presence or absence of 48 h of Trolox treatment. RD, regulatory domain. **b**, H2Aub-ChIP for Atg7<sup>WT</sup> and Atg7<sup>ΔPax7</sup> satellite cells treated as in **a**. **c**, Western blotting for cells in **b**; for full gel scan see Supplementary Fig. 1. Data show mean ± s.e.m. Comparisons by two-sided Mann-Whitney *U*-test. *P* values indicated. Sample numbers were *n* = 3 animals per group for **a**–**c**.

Loss of the polycomb repressive complex-1 (PRC1)-mediated H2A monoubiquitination of lysine 119 (H2Aub) at *INK4a* (also known as *Cdkn2a*) locus drives p16<sup>INK4a</sup> induction in geriatric satellite cells<sup>8</sup> (Extended Data Fig. 5i). We found that Trolox treatment restored *INK4a* locus H2Aub modification in geriatric and Atg7-deficient satellite cells (Fig. 6a, b), resulting in p16<sup>INK4a</sup> repression, and this reduced senescence while promoting proliferation (Figs 5c–f and 6c and Extended Data Fig. 5j). Genetic silencing (with short-hairpin RNA) of *INK4a* restored proliferation in Atg7<sup>ΔPax7</sup> satellite cells while reducing the expression of senescence-associated genes and the number of SA-β-gal<sup>+</sup> cells, and augmenting their regenerative capacity (Extended Data Figs 5k and 6a, b). Thus, the ROS-induced p16<sup>INK4a</sup> axis links impaired autophagy and senescence in ageing satellite cells.

## Defective autophagy in aged human cells

Skeletal muscles from geriatric individuals show sarcopenia and presence of senescent satellite cells (Extended Data Fig. 7a, b)<sup>8</sup>. As in mice, human satellite cells from geriatric individuals showed defective protein and organelle clearance, as indicated by p62 and mitochondrial accumulation (Extended Data Fig. 7c, d) compared to young cells, which was tightly associated with increased ROS levels (Extended Data Fig. 7d, e) and SA-β-gal<sup>+</sup> cells (Extended Data Fig. 7f), consistent with reduced proliferative potential (Extended Data Fig. 7g). The causal role of impaired autophagy on the gerocomversion of ageing human satellite cells under proliferative pressure was supported by the capacity of rapamycin to revert the abnormal mitochondrial content, protein aggregates and ROS (Extended Data Fig. 7c, d), and senescence phenotype (Extended Data Fig. 7f–i). Thus, restoration of autophagy and organelle homeostasis in aged human satellite cells suffices to rescue senescence, as is the case in murine satellite cells.

## Discussion

In tissues with little turnover, reversible quiescence is the normal stem-cell state throughout life. However, quiescence is known to be progressively lost with ageing due to systemic/niche- and intrinsic-factor alterations<sup>2,7</sup>. Recent studies showed that at geriatric age, the normal stem-cell quiescent state is substituted by an irreversible senescence state, which results in a numerical and functional decline of stem cells<sup>8</sup>. The mechanisms accounting for the maintenance of quiescence, preservation of the stem-cell pool and prevention of senescence during an individual's life remain unknown. Our results demonstrate that quiescent satellite cells are equipped with cytoprotective and cellular quality-control mechanisms that actively repress the senescence program, thereby preserving the integrity and fitness of cells. We provide evidence of loss of autophagy in satellite cells that occurs with ageing,



resulting in accumulation of damaged proteins and organelles, leading to senescence and stem-cell exhaustion. Consistent with this finding, genetic inhibition of autophagy specifically in satellite cells of young mice caused rapid entry into senescence, resulting in numerical and functional exhaustion of stem cells, and defective muscle regeneration. These findings were surprising, considering that a decline in basal autophagy in quiescent stem cells of physiologically aged mammalian organisms has not been described before. Autophagy is usually considered to be an effector pathway, rather than a cause, of senescence, particularly in oncogene-induced senescence<sup>30–33</sup>.

How autophagy balances quiescence and senescence in muscle stem cells is unknown. Here we show that in adult resting muscle, quiescent stem cells attenuate proteotoxicity by maintaining a high basal autophagy flux, constituting a homeostatic ‘clean up’ process. This function is particularly critical in non-dividing stem cells, in which mitotic dilution of intracellular toxic debris does not take place<sup>20,34</sup>. Autophagy failure in aged resting stem cells leads to accumulation of damaged proteins and dysfunctional organelles, specially mitochondria, which generates enhanced ROS levels that cause DNA damage and senescence entry, consistent with previous studies<sup>35–42</sup>. Indeed, we uncover ROS as a key epigenetic regulator of the senescence-promoting gene *INK4a* in ageing stem cells, by impeding PRC1-mediated lysine 119 H2A ubiquitination, the required epigenetic mark for *INK4a* locus silencing. Consistent with this, treatment of geriatric mice (and mice with satellite-cell-specific *Atg7* deficiency) with antioxidants not only restored PRC1-mediated *INK4a* locus repression and prevented satellite cell senescence, but also restored regenerative capacity. Signs of impaired autophagy and loss of proteostasis, correlating with senescence and defective myogenic functions, were also observed in human satellite cells from geriatric individuals.

At variance with our findings, a recent study demonstrated that, upon *in vitro* stress, autophagy does not decline, but is even induced in haematopoietic stem cells with ageing, consistent with maintenance of haematopoietic stem cell number<sup>43</sup>. Thus, we propose that long-lived quiescent stem cells within low turnover tissues primarily rely on autophagy to preserve fitness and avoid senescence, and that stem cells of skeletal muscle in particular lose this protection during ageing (Extended Data Fig. 7j). Notably, a recent study also reported that autophagy is needed for the activation of young satellite cells<sup>44</sup>. Furthermore, in the whole musculature, age-associated myofibre degeneration and mitochondrial dysfunction could also be alleviated by autophagy reactivation<sup>19,45</sup>.

Our studies thus demonstrate that autophagy is a decisive factor in the switch between the quiescence and senescence fate of muscle stem cells (Extended Data Fig. 7j). Although ageing-induced senescence is often viewed as an inescapable and irremediable process, we provide evidence that *in vivo* restoration of constitutive autophagy (or neutralization of excessive ROS) averts intracellular damage accumulation, and prevents satellite cell senescence and functional decline in old mice, as well as in aged human stem cells, reinforcing the notion that the intrinsic-ageing clock in stem cells can be pharmacologically manipulated.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 21 November 2014; accepted 29 October 2015.**

- Cheung, T. H. & Rando, T. A. Molecular regulation of stem cell quiescence. *Nature Rev. Mol. Cell Biol.* **14**, 329–340 (2013).
- Comai, G. & Tajbakhsh, S. Molecular and cellular regulation of skeletal myogenesis. *Curr. Top. Dev. Biol.* **110**, 1–73 (2014).
- Yin, H., Price, F. & Rudnicki, M. A. Satellite cells and the muscle stem cell niche. *Physiol. Rev.* **93**, 23–67 (2013).
- Montarras, D., L'Honore, A. & Buckingham, M. Lying low but ready for action: the quiescent muscle satellite cell. *FEBS J.* **280**, 4036–4050 (2013).
- Grounds, M. D. Therapies for sarcopenia and regeneration of old skeletal muscles: more a case of old tissue architecture than old stem cells. *Bioarchitecture.* **4**, 81–87 (2014).
- García-Prat, L., Sousa-Victor, P. & Muñoz-Cánoves, P. Functional dysregulation of stem cells during aging: a focus on skeletal muscle stem cells. *FEBS J.* **280**, 4051–4062 (2013).
- Chakkalakal, J. V., Jones, K. M., Basson, M. A. & Brack, A. S. The aged niche disrupts muscle stem cell quiescence. *Nature* **490**, 355–360 (2012).
- Sousa-Victor, P. et al. Geriatric muscle stem cells switch reversible quiescence into senescence. *Nature* **506**, 316–321 (2014).
- Sousa-Victor, P., García-Prat, L., Serrano, A. L., Perdiguer, E. & Muñoz-Cánoves, P. Muscle stem cell aging: regulation and rejuvenation. *Trends Endocrinol. Metab.* **26**, 287–296 (2015).
- Cosgrove, B. D. et al. Rejuvenation of the muscle stem cell population restores strength to injured aged muscles. *Nature Med.* **20**, 255–264 (2014).
- Bernet, J. D. et al. p38 MAPK signaling underlies a cell-autonomous loss of stem cell self-renewal in skeletal muscle of aged mice. *Nature Med.* **20**, 265–271 (2014).
- Price, F. D. et al. Inhibition of JAK–STAT signaling stimulates adult satellite cell function. *Nature Med.* **20**, 1174–1181 (2014).
- Tierney, M. T. et al. STAT3 signaling controls satellite cell expansion and skeletal muscle repair. *Nature Med.* **20**, 1182–1186 (2014).
- Cuervo, A. M. et al. Autophagy and aging: the importance of maintaining “clean” cells. *Autophagy* **1**, 131–140 (2005).
- He, C. & Klionsky, D. J. Regulation mechanisms and signaling pathways of autophagy. *Annu. Rev. Genet.* **43**, 67–93 (2009).
- Fukada, S. et al. Molecular signature of quiescent satellite cells in adult skeletal muscle. *Stem Cells* **25**, 2448–2459 (2007).
- Liu, L. et al. Chromatin modifications as determinants of muscle stem cell quiescence and chronological aging. *Cell Rep.* **4**, 189–204 (2013).
- Pallafacchina, G. et al. An adult tissue-specific stem cell in its niche: a gene profiling analysis of *in vivo* quiescent and activated muscle satellite cells. *Stem Cell Res.* **4**, 77–91 (2010).
- Carnio, S. et al. Autophagy impairment in muscle induces neuromuscular junction degeneration and precocious aging. *Cell Rep.* **8**, 1509–1521 (2014).
- Rubinsztein, D. C., Marino, G. & Kroemer, G. Autophagy and aging. *Cell* **146**, 682–695 (2011).
- Mizushima, N., Yamamoto, A., Matsui, M., Yoshimori, T. & Ohsumi, Y. *In vivo* analysis of autophagy in response to nutrient starvation using transgenic mice expressing a fluorescent autophagosome marker. *Mol. Biol. Cell* **15**, 1101–1111 (2004).
- Klionsky, D. J. et al. Guidelines for the use and interpretation of assays for monitoring autophagy in higher eukaryotes. *Autophagy* **4**, 151–175 (2008).
- Zhu, J., Dagda, R. K. & Chu, C. T. Monitoring mitophagy in neuronal cell cultures. *Methods Mol. Biol.* **793**, 325–339 (2011).
- Mammucari, C. et al. FoxO3 controls autophagy in skeletal muscle *in vivo*. *Cell Metab.* **6**, 458–471 (2007).
- Morselli, E. et al. Spermidine and resveratrol induce autophagy by distinct pathways converging on the acetylproteome. *J. Cell Biol.* **192**, 615–629 (2011).
- van Deursen, J. M. The role of senescent cells in ageing. *Nature* **509**, 439–446 (2014).
- Kimura, S., Noda, T. & Yoshimori, T. Dissection of the autophagosome maturation process by a novel reporter protein, tandem fluorescent-tagged LC3. *Autophagy* **3**, 452–460 (2007).
- Muñoz-Espín, D. & Serrano, M. Cellular senescence: from physiology to pathology. *Nature Rev. Mol. Cell Biol.* **15**, 482–496 (2014).
- Blagosklonny, M. V. Selective anti-cancer agents as anti-aging drugs. *Cancer Biol. Ther.* **14**, 1092–1097 (2013).
- Young, A. R. et al. Autophagy mediates the mitotic senescence transition. *Genes Dev.* **23**, 798–803 (2009).
- Narita, M. et al. Spatial coupling of mTOR and autophagy augments secretory phenotypes. *Science* **332**, 966–970 (2011).
- Pérez-Mancera, P. A., Young, A. R. & Narita, M. Inside and out: the activities of senescence in cancer. *Nature Rev. Cancer* **14**, 547–558 (2014).
- Capparelli, C. et al. Autophagy and senescence in cancer-associated fibroblasts metabolically supports tumor growth and metastasis via glycolysis and ketone production. *Cell Cycle* **11**, 2285–2302 (2012).
- Flach, J. et al. Replication stress is a potent driver of functional decline in ageing haematopoietic stem cells. *Nature* **512**, 198–202 (2014).
- Kodama, R. et al. ROS-generating oxidases Nox1 and Nox4 contribute to oncogenic Ras-induced premature senescence. *Genes Cells* **18**, 32–41 (2013).
- Ramsey, M. R. & Sharpless, N. E. ROS as a tumour suppressor? *Nature Cell Biol.* **8**, 1213–1215 (2006).
- Ito, K. et al. Reactive oxygen species act through p38 MAPK to limit the lifespan of hematopoietic stem cells. *Nature Med.* **12**, 446–451 (2006).
- Lee, A. C. et al. Ras proteins induce senescence by altering the intracellular levels of reactive oxygen species. *J. Biol. Chem.* **274**, 7936–7940 (1999).
- Mandal, P. K., Blanpain, C. & Rossi, D. J. DNA damage response in adult stem cells: pathways and consequences. *Nature Rev. Mol. Cell Biol.* **12**, 198–202 (2011).
- Shao, L. et al. Reactive oxygen species and hematopoietic stem cell senescence. *Int. J. Hematol.* **94**, 24–32 (2011).
- Lerner, C. et al. Reduced mammalian target of rapamycin activity facilitates mitochondrial retrograde signaling and increases life span in normal human fibroblasts. *Aging Cell* **12**, 966–977 (2013).
- López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1217 (2013).
- Warr, M. R. et al. FOXO3A directs a protective autophagy program in haematopoietic stem cells. *Nature* **494**, 323–327 (2013).

44. Tang, A. H. & Rando, T. A. Induction of autophagy supports the bioenergetic demands of quiescent muscle stem cell activation. *EMBO J.* **33**, 2782–2797 (2014).
45. Lee, J. H. *et al.* Sestrin as a feedback inhibitor of TOR that prevents age-related pathologies. *Science* **327**, 1223–1228 (2010).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We are indebted to G. Mariño for the gift of GFP–LC3 transgenic mice, C. Keller and M. Capecchi for Pax–Cre mouse lines, J. Ruberte for TEM studies help, E. Masliah and K. Kosberg for Atg7 lentivirus; M. Raya, M. Jardí, and V. Lukesova for their technical contributions, and especially J. Guerra for help in microarray experiments and P. Sousa-Victor for initial findings; J. Martín-Caballero (PRBB Animal Facility) and O. Fornas (UPF/CRG FACS Facility) for technical help, and the KS Society. The authors acknowledge funding from MINECO, Spain (SAF2012-38547, SAF2015-67369-R, PLE2009-0124; SAF2009-08374; “María de Maeztu” Programme for Units of Excellence in R&D MDM-2014-0370), AFM, E-Rare/ERANET, Fundació Marató TV3, MDA, EU-FP7 (Myoage, Optistem and Endostem) and DuchennePP-NL. M.M.-V. acknowledges funding from ISCIII, Spain (FIS-PS09/01267, FIS-PI13/02512, CP09/00184, PI14/01529) and CIBERNED; and MS from the European Union

ERC (282310-MyoPHAGY) and Foundation Leducq. L.G.-P. was supported by a Predoctoral Fellowship from Programa de Formación de Personal Investigador (Spain).

**Author Contributions** L.G.-P. designed and performed most experiments, analysed data, interpreted results and wrote the manuscript. A.L.S. and E.P. designed and performed experiments, and helped in interpreting results and editing the manuscript. M.M.-V. and M.S. helped in designing and interpreting some experiments and results and editing the manuscript. L.O., V.R.-B. and S.G. performed some experiments and provided technical support. E.R. provided technical support in microscopy. J.R.-U. and E.B. performed ChIP experiments and helped in interpreting results. P.M.-C. conceived the project, designed experiments, interpreted results and wrote the manuscript.

**Author Information** Microarray data have been deposited into the NCBI Gene Expression Omnibus under accession number GSE70376. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.M.-C. ([pura.munoz@upf.edu](mailto:pura.munoz@upf.edu)).



## METHODS

**Data reporting.** No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment.

**Mice.** Male mice (C57BL/6 (wild-type, WT), LC3-GFP, the offspring of intercrossing *Atg7<sup>fl/fl</sup>* with *Pax7<sup>Cre</sup>* and *Pax7<sup>CreER</sup>* lines) were used at different ages. GFP-LC3 mice were provided by G. Mariño. Mice with the *Atg7* gene deletion in satellite cells, as an inducible or constitutive deletion, were generated by breeding *Atg7<sup>fl/fl</sup>* mice (previously described in ref. 46) with the *Pax7<sup>Cre</sup>* and *Pax7<sup>CreER</sup>* lines (provided by C. Keller and M. Capecchi, respectively). All animal experiments were approved by the ethics committee of the (Barcelona Biomedical Research Park (PRBB) and by the Catalan Government and used sex-, age- and weight-matched littermate animals.

When needed, Cre activity was induced by intraperitoneal injection (one injection per day for 4 days) with 5 mg per 25 g body weight of tamoxifen (Sigma; 10 mg ml<sup>-1</sup> in corn oil).

**Induction of muscle regeneration.** Mice were anaesthetized with ketamine and xylazine (80:10 mg kg<sup>-1</sup>, intraperitoneally). Regeneration of skeletal muscle was induced by intramuscular injection of cardiotoxin (CTX, Latoxan; 10<sup>-5</sup> M) in the tibialis anterior muscle of the mice as described<sup>47</sup>. At the indicated times after injury, mice were euthanized and muscles were dissected, frozen in isopentane cooled with liquid nitrogen, and stored at -80°C until analysis. For GFP immunostaining of samples, muscles were prefixed for 2 h in 2% paraformaldehyde at 4°C, and were embedded in 15% sucrose overnight at 4°C and then frozen in isopentane cooled with liquid nitrogen.

**Satellite cell isolation by FACS.** Muscles were mechanically disaggregated and incubated in Ham's F10 media containing 0.8% collagenase D (Roche) and 0.125% trypsin and EDTA at 37°C with agitation, for 25 min and the supernatant was then filtered. The digestion procedure was repeated four times and the supernatants were collected. Cells were incubated in lysis buffer (BD Pharm Lyse) for 10 min on ice, re-suspended in PBS with 2.5% goat serum and counted. PE-Cy7-conjugated anti-CD31 (Biolegend 102418), anti-CD11b (Biolegend 101215/16) and anti-Sca-1 (Biolegend 108113/14) antibodies were used to exclude the Lin (-) negative population and Alexa647-conjugated anti-CD34 (BD Pharmigen 560230) and PE-conjugated anti-α7-integrin (Ablab AB10STMW215) were used for double-positive staining of quiescent satellite cells. Cells were sorted using a FACS Aria II (BD). Isolated satellite cells were used either for RNA extraction or were cultured in Ham's F10 supplemented with 30% FBS and bFGF (0.025 µg ml<sup>-1</sup>) (growth medium) for proliferation assays or plated on glass slides (Thermo Scientific 177402) for immunostaining analysis.

**Flow cytometry analysis.** FACS isolated satellite cells (see above) were stained with different dyes for flow cytometry analysis. Staining for mitochondria, lysosomes and ROS was performed by incubating cells at 37°C with 1 µM tetramethylrhodamine, methyl ester TMRM (T-668), 100 nM MitoTracker Green FM (M7514), 100 nM MitoTracker Red CMXRos (M7512), 500 nM LysoTracker Green DND-26 (L7526) and 5 µM CellROX Green reagent (C10444), following the manufacturer's protocols (Invitrogen) and directly analysed without fixing. Cell analysis was performed in FACS LSR Fortessa (Becton Dickinson). For MFI determination, we used the flow cytometry analysis software Flowlogic. MFI refers to the fluorescence intensity of each event (on average) of the selected cell population, in the chosen fluorescence channel.

**Whole-transcriptome analysis of FACS-sorted satellite cells.** FACS-sorted satellite cells were collected in lysis buffer and RNA extraction was performed using RNeasy Micro kit (Qiagen). The cDNA was used for transcriptome analysis by Agilent SurePrint G3 Mouse GE 8 × 60 K high density microarray slides, performed at the microarray Unit of CRG (Barcelona, Spain). Microarray analysis was performed with 3 animals each. Data was normalized using cyclic loess, and differentially expressed genes were identified using AFM 4.0 (ref. 48) for all pairwise comparisons. Raw data was taken from the Feature Extraction output files and was corrected for background noise using the normexp method. To assure comparability across samples quantile normalization was used. Differential expression analysis was carried out on non-control probes with an empirical Bayes approach on linear models (limma). Results were corrected for multiple testing according to the false discovery rate (FDR) method. Statistical analysis was performed with the Bioconductor project (<http://www.bioconductor.org/>) in the R statistical environment. Venn diagrams were generated using BioVenn<sup>49</sup>.

**In vivo treatments.** Autophagy of aged C57BL/6 and GFP-LC3 mice was induced as follows, one group of mice was injected i.p. with 4 mg per kg body weight of rapamycin (LC Laboratories) or vehicle (DMSO) every other day for 2 weeks; a second group was injected i.p. with 30 mg per kg body weight of Trolox (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid, Sigma) or

vehicle (DMSO) daily for 2 weeks; and the third group of mice was treated with 3 mM spermidine (S2626 Sigma) in drinking water for 2 weeks.

**Satellite cell engraftment.** Satellite cell transplants were performed as in ref. 8, following an adapted protocol<sup>50</sup>. Quiescent FACS-isolated satellite cells were collected, re-suspended in 20% FBS Ham's F10 medium and injected into muscles of recipient mice previously injured with cardiotoxin the day before. The recipient mice were SCID mice. For each mouse, 10,000 cells were injected. At 4 days (for proliferation, senescence analyses) or 1 month (muscle regeneration) after cell injections, engrafted muscles were collected and processed for muscle histology. Results are expressed as relative number of GFP<sup>+</sup> per muscle section, with respect to the control data for young cells, which was set at 100%.

**In vitro treatments.** Experiments for *in vitro* rescue of defective autophagy in satellite cells were performed in 20% FBS containing Ham's F10 medium (growth medium), and with the addition of either rapamycin (100 ng ml<sup>-1</sup>, LC Laboratories), Trolox (100 µM, Sigma), spermidine (5 µM, Sigma) or vehicle (DMSO) for 48 h. Mitochondrial, lysosomal, and ROS analyses or ChIP experiments were performed immediately after treatments, whereas proliferation assays (BrdU staining) and senescence analysis (SA-β-gal assay and determination of RNA and protein expression of senescence markers), were performed 96 h after treatments.

For the satellite cell treatments for *in vivo* engraftment in injured muscles, fresh FACS-isolated satellite cells from resting muscle of young and geriatric mice were treated for 48 h with rapamycin (100 ng ml<sup>-1</sup>, LC Laboratories), Trolox (100 µM, Sigma) or vehicle (DMSO) before engraftment into pre-injured muscles of recipient mice. For each mouse, 10,000 cells were injected. At 4 days after cell injections, engrafted muscles were collected and processed for muscle histology.

Bafilomycin (10 nM Sigma B1793) was used to block autophagy for 4 h at 37°C and to analyse autophagosome accumulation by FACS, immunostaining and western blotting. CCCP (carbonyl cyanide 3-chlorophenylhydrazone, 10 µM Sigma C2759), which abolishes the link between the respiratory chain and the phosphorylation system in intact mitochondria, causes mitochondria uncoupling and was used to treat satellite cells *in vitro* for 1 h to induce the selective autophagy of CCCP-damaged mitochondria (mitophagy).

**Plasmid transfection.** Freshly isolated cells were transfected with mRFP-GFP-LC3 (ref. 23) plasmid using Lipofectamine 3000 (Invitrogen), and treated for 48 h with Trolox (25 µl ml<sup>-1</sup>, Sigma) or vehicle (DMSO) and analysed on glass slides (Thermo Scientific 177402). Cells were fixed with 4% paraformaldehyde in PBS for 10 min and the nuclei were stained with DAPI (Invitrogen). After washing, glass slides were mounted with Mowiol. Measuring autophagy flux through this method is based on the concept of lysosomal quenching of GFP. GFP is a stably folded protein and relatively resistant to lysosomal proteases. However, the low pH inside the lysosome quenches the fluorescent signal of GFP, which makes it difficult to trace the delivery of GFP-LC3 to lysosomes. In contrast, RFP exhibits more stable fluorescence in acidic compartments, and mRFP-LC3 can be readily detected in autolysosomes. By exploiting the difference in the nature of these two fluorescent proteins (that is, lysosomal quenching of GFP fluorescence versus lysosomal stability of RFP fluorescence), autophagic flux can be morphologically traced with an mRFP-GFP-LC3 tandem construct<sup>23</sup>. With this tandem construct, autophagosomes and autolysosomes are labelled with yellow (mRFP and GFP) and red (mRFP only) signals, respectively.

**Proliferation assay.** Satellite cells were labelled with BrdU (1.5 µg ml<sup>-1</sup>; Sigma) for 1 h. BrdU-labelled cells were detected by immunostaining using rat anti-BrdU antibody (Oxford Biotechnology; 1:500) and a specific secondary biotinylated goat anti-rat antibody (Jackson ImmunoResearch; 1:250). Antibody binding was visualized using Vectastain Elite ABC reagent (Vector Laboratories) and DAB. BrdU-positive cells were quantified as the percentage of the total number of cells analysed.

**SA-β-gal activity.** SA-β-gal activity was detected in satellite cells using the senescence β-galactosidase staining kit (Cell signaling), according to the manufacturer's instructions. SA-β-gal<sup>+</sup> cells were quantified as percentage of the total number of cells analysed.

**Lentivirus infection.** Freshly isolated satellite cells were *ex vivo* infected with distinct lentivirus for 12 h. Medium was replaced and cells were transplanted into injured muscle of recipient mice for *in vivo* analysis, or subjected to *in vitro* assays. LV-Atg7, used for Atg7 overexpression in satellite cells, was provided by Eliezer Masliah's laboratory<sup>51</sup>. LV-sh-p16<sup>INK4a</sup>, used to silence *INK4a*, and LV-sh-scramble (used as control), were previously described in ref. 8.

**Heterografting experiments.** Extensor digitorum longus (EDL) muscles from geriatric wild-type mice were infected with lentivirus (LV-Atg7 or LV-GFP, as well as LV-sh-p16<sup>INK4a</sup> or LV-sh-scramble) and grafted immediately onto the tibialis anterior muscle of young wild-type recipient mice, and regeneration

(formation of new myofibres derived from EDL-associated satellite cells) in the transplanted EDL muscles was analysed after 6 or 8 days. Fibre size of eMHC<sup>+</sup> myofibre was analysed using the Fiji program.

**RT-qPCR: RNA extraction, cDNA synthesis and PCR.** Total RNA was isolated from either FACS-isolated satellite cells of mouse muscle tissue or human myoblasts obtained from human muscle biopsies, using Tripure reagent (Roche Diagnostic Corporation) or RNeasy Micro kit (Qiagen), and analysed by RT-qPCR. For qPCR experiments, DNase digestion of 10 mg of RNA was performed using 2 U DNase (Turbo DNA-free, Ambion). Complementary DNA (cDNA) was synthesized from total RNA using the First-Strand cDNA Synthesis kit (Amersham Biosciences). Real-time PCR reactions were performed on a LightCycler 480 System using Light Cycler 480 SYBR Green I Master reaction mix (Roche Diagnostic Corporation) and specific primers. Thermocycling conditions were as follows: initial step of 10 min at 95 °C, then 50 cycles of 15 s denaturation at 94 °C, 10 s annealing at 60 °C and 15 s extension at 72 °C. Reactions were run in triplicate, and automatically detected threshold cycle values were compared between samples. Transcript of the ribosomal protein L7 housekeeping gene was used as endogenous control, with each unknown sample normalized to L7 content. The following primers were used, *INK4a*, forward: CATCTGGAGCAGCATGGAGTC, reverse: GGGTACGACCGAAAGAGTTCG; *p21<sup>CIP1</sup>* (also known as *Cdkn1a*) forward: CCAGGCCAA GATGGTGTCTT, reverse: TGAGAAAGGATCA GCCATTGC; *MyoD* (also known as *Myod1*), forward: GCCCGCTGAGC AAAGTGAATG, reverse: CAGCGGTCCAGGTGCGTAGAAG; *Mgn* (*Myog*), forward, GGTGTGTAAGAGGAAGTCTGTG, reverse: TAGGCGCTCAAT GTACTGGAT; *Ki67* (*Mki67*), forward, ACCGTGGAGTAGTTTATCTGGG, reverse, TGTTTCCAGTCCGCTT-ACTTCT; *p15<sup>INK4b</sup>* (also known as *Cdkn2b*), forward, TCTTGCATCTCCACAGCTG, reverse, TCCAGGTTTCCCA TTTAGC; *Atg7*, forward, TCTGGGAAGCCATAAAGTCAGG, reverse, GCGAAGTCCAGGAGCAGAA.

**Electron microscopy images.** For electron microscopy images, tibialis anterior muscles from 3- and 24-month-old wild-type mice were fixed with 2% paraformaldehyde and 2.5% glutaraldehyde in phosphate buffer (0.1 M, pH 7.4). Samples were processed by the CCit Microscopy Facility at the University of Barcelona. Images were acquired using a Jeol 1010 microscope, working at 80 kV and equipped with a CCD Megaview III camera. Identification of satellite cells in skeletal muscle by electron microscopy was based on cell size, content of heterochromatin and position with respect to basal lamina.

**Western blotting.** Preparation of mouse and human satellite cell lysates and western blotting was performed as described previously in ref. 52. Antibodies used were: anti-p62/SQSTM1 antibody produced in rabbit (Sigma P0067), rabbit anti-LC3 (Novus Biologicals NB100-2331), phospho-S6 ribosomal protein (Ser240/244) XP rabbit monoclonal antibody (Cell Signaling 5364), rabbit anti-p16 (Santa Cruz Biotechnology sc-1207), rabbit anti-parkin (Abcam ab15954), S6 ribosomal protein (54D2) mouse (Cell Signaling 2317),  $\gamma$ H2AX Ser 139 (Cell Signaling 2577S), rabbit anti-53BP1 (Abcam ab21083) and Tubulin (Sigma T-6199).

**ChIP.** Briefly, freshly isolated satellite cells were cultured with Trolox or vehicle (DMSO) for 48 h and crosslinked with 1% formaldehyde for 15 min at room temperature. For each ChIP, 300,000 cells were lysed in 130  $\mu$ l of lysis buffer B (Low Cell ChIP Kit, Diagenode) and chromatin was sonicated for 10 min in a M220 Focused-ultrasonicator, Covaris (Duty cycle 5%, Peak incident power 75 W and 200 cycles per burst). Sonicated chromatin was then diluted and subjected to immunoprecipitation with 3  $\mu$ l of antibody against ubiquitinated histone (Ubiquitinyl-Histone) H2A (Lys119) (D27C4) (Cell Signaling, 8240) or 3  $\mu$ l of IgG. Bound fraction and input were analysed by qPCR using specific primer sets for the *INK4a* locus. *INK4a*\_RD forward, GGTCTCCCTAGCAGGATTC, reverse GCCTGTCATTAAACAGGGTGA; *INK4a*\_exon1 forward, CCGGAGCCACCCATTAACTA, reverse CAAAGACTT CTCAAAAATAAGACACTGAAA; *INK4a*\_exon2 forward, CCCAACACCC ACTTGAGGAA, reverse, CAGAGGTCACAGGCATCGAA.

**Histology and immunohistochemistry in muscle cryosections.** Tibialis anterior and extensor digitorum longus (EDL) muscles were frozen in isopentane cooled with liquid nitrogen, and stored at -80 °C until analysis. Then 10- $\mu$ m sections were collected from muscles and were either stained with haematoxylin and eosin or immunostained. Labelling of cryosections with mouse monoclonal primary antibodies was performed using the peroxidase or fluorescein M.O.M. kit staining (Vector Laboratories) according to the manufacturer's instructions. Double immunostaining was performed by sequential addition of each primary and secondary antibody using appropriate positive and negative controls. Sections were air dried, fixed on 2–4% paraformaldehyde, washed on PBS and incubated with primary antibodies according to manufacturer's instructions after blocking for 1 h at room temperature with

a high-protein-containing solution in PBS (Vector Laboratories). The slides were then washed with PBS and incubated with the appropriate secondary antibodies and labelling dyes. For immunofluorescence, secondary antibodies were coupled to Alexa-488, Alexa-568 or Alexa-647 fluorochromes, and nuclei were stained with DAPI (Invitrogen). After washing, tissue sections were mounted with Mowiol.

**Antibodies used for immunohistochemistry.** Immunohistochemistry on muscle cryosections or isolated satellite cells was performed with the following antibodies: GFP (Invitrogen A6455 and Aves labs GFP-1020), anti-eMHC (F1.652), anti-Pax7 (DSHB), p16 (Santa Cruz sc-1207),  $\gamma$ H2AX Ser139 (2577S), rabbit polyclonal anti-MyoD (Santa Cruz Biotechnology sc-760), anti-myogenin (DSHB F5D), poly-ubiquitinated proteins, multi-ubiquitin chains, mouse monoclonal antibody (Enzo life sciences PW8805), anti-p62/SQSTM1 antibody produced in rabbit (Sigma P0067), mouse monoclonal antibody to LC3 (NanoTools 5F10), LAMP-1 (Santa Cruz Biotechnology sc-19992), phospho-S6 ribosomal protein (Ser240/244) XP rabbit monoclonal antibody (Cell Signaling 5364), anti-CD56 (BD Pharmingen 556325), anti-TOM20 (ab56783).

**Human muscle samples.** Muscle biopsies from 8 adults and 10 geriatric (28  $\pm$  7 and 83  $\pm$  7 years old, respectively) human subjects were obtained via the Tissue Banks for Research from Vall d'Hebron and Sant Joan de Deu Hospitals and especially via the EU/FP7 Myoage Consortium. Muscle biopsies were taken from the vastus lateralis muscle under local anaesthesia (2% lidocaine). A portion of the muscle tissue was directly frozen in melting isopentane and stored at -80 °C until analysis. Human primary myoblasts from 5 young/adult (25  $\pm$  4 years old) and 5 geriatric (75  $\pm$  4 years old) subjects were obtained from the EU/FP7 Myoage Consortium or purchased from Cook Myosite and cultured following the provided instructions.

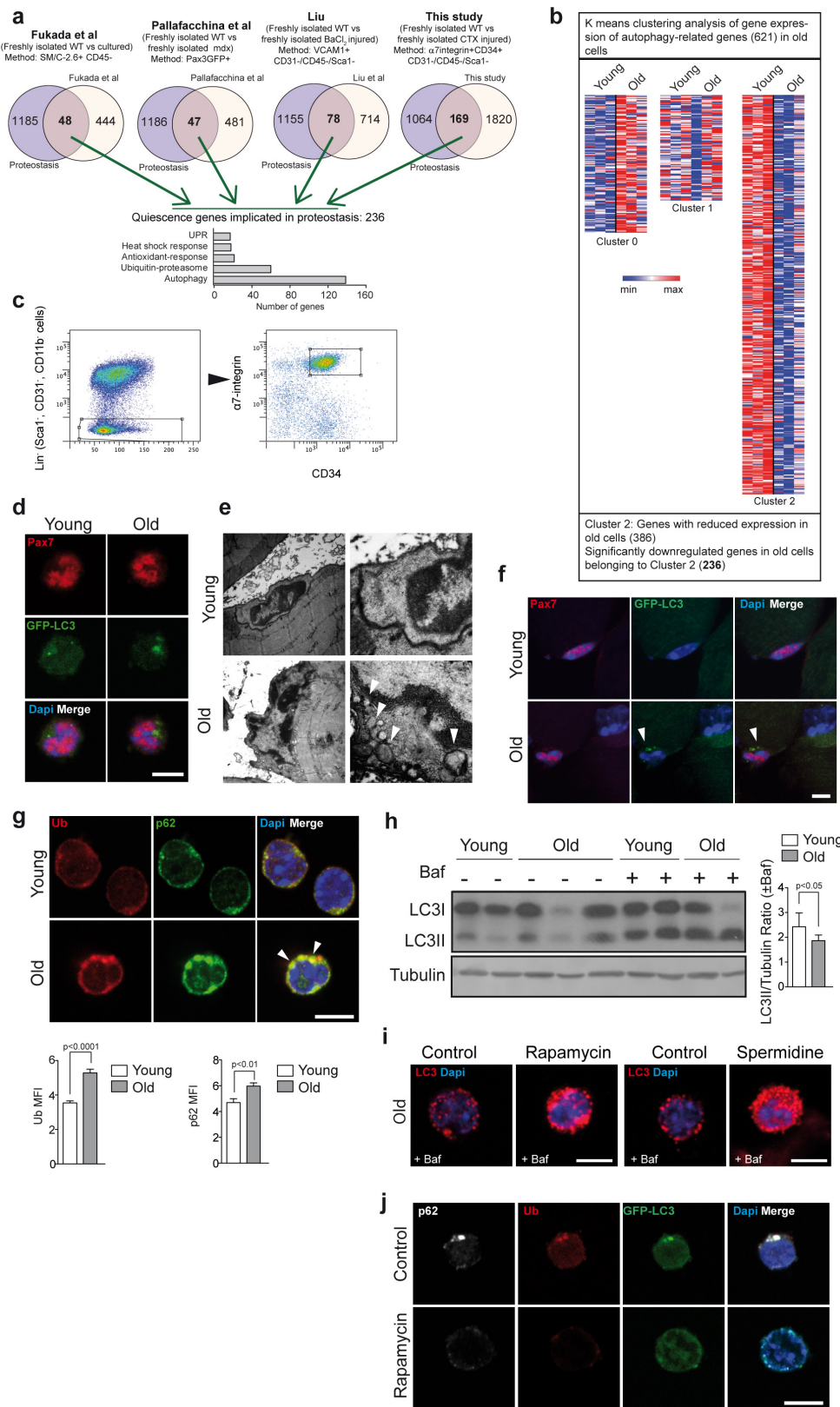
**Digital image acquisition and processing.** Digital images were acquired using: (1) an upright microscope DMR6000B (Leica) equipped with a DFC300FX camera for immunohistochemical colour pictures and a Hamamatsu ORCA-ER camera for immunofluorescence pictures; (2) confocal images of muscle sections or isolated satellite cells were taken using either a Zeiss LSM-780 confocal system with a Plan-Apochromat 63  $\times$  /1.4 NA oil objective or a Leica SPE confocal laser scanning microscope system with HCX PL Fluotar 10  $\times$  /0.30 NA, 20  $\times$  /0.50 NA and 40  $\times$  /0.75 NA objectives. The different fluorophores (3 to 4) were excited using the 405, 488, 568 and 633 nm excitation lines. Acquisition was performed using Zeiss LSM software Zen Black or Leica Application or LAS AF software (Leica). Images were composed and edited in Photoshop CS5 (Adobe), in which background was reduced using brightness and contrast adjustments applied to the whole image. To assess myofibre size, individual fibres were manually outlined and their cross-sectional area (CSA) was determined with the public domain image analysis software Fiji. Fluorescence intensity of selected proteins for each cell was quantified using Fiji software and the average of relative fluorescence was expressed as MFI.

The number and percentage of cellular area occupied by GFP-LC3 puncta were determined on digital images with Fiji and the cell image analysis software CellProfiler<sup>53</sup>. Co-localization of RFP-LC3 and GFP-LC3 puncta was determined on the maximum projection of three z-sections using a Fiji automated macro pipeline calculating single and double-positive autophagosomes. Co-localization of p62 and ubiquitin was determined on digital images Fiji, according to ref. 54, with respect to the total cellular area. The Pearson's coefficient ( $r$ ) was used to analyse the correlation of the intensity values of green and red pixels in dual-channel images. This coefficient measures the strength of the linear relationship between the intensities in two images calculated by linear regression and ranges from 1 to  $\times$ 1, with 1 standing for complete positive correlation and  $\times$ 1 for a negative correlation, with zero standing for no correlation<sup>51</sup>. Video reconstructions of autophagosomes were generated in Imaris software using full confocal z-stacks (around 20) of each cell. The z-stacks were previously imported to Fiji software for background adjustments and then deconvolved using the blind-deconvolution wizard in Huygens software.

**Statistical analysis.** For mouse experiments, no specific blinding method was used, but mice in each sample group were selected randomly. The sample size ( $n$ ) of each experimental group is described in each corresponding figure legend, and all experiments were repeated at least with three biological replicates. GraphPad Prism software was used for all statistical analyses. Quantitative data displayed as histograms are expressed as means  $\pm$  standard error of the mean (represented as error bars). Results from each group were averaged and used to calculate descriptive statistics. Mann-Whitney  $U$ -test (independent samples, two-sided) was used for pairwise comparisons among groups at each time point. Statistical significance was set at a  $P < 0.05$ .



46. Masiero, E. *et al.* Autophagy is required to maintain muscle mass. *Cell Metab.* **10**, 507–515 (2009).
47. Suelves, M. *et al.* uPA deficiency exacerbates muscular dystrophy in MDX mice. *J. Cell Biol.* **178**, 1039–1051 (2007).
48. Breitkreutz, B. J., Jorgensen, P., Breitkreutz, A. & Tyers, M. AFM 4.0: a toolbox for DNA microarray analysis. *Genome Biol.* **2**, <http://dx.doi.org/10.1186/gb-2001-2-8-software0001> (2001).
49. Hulsen, T., de Vlieg, J. & Alkema, W. BioVenn—a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* **9**, 488 (2008).
50. Sacco, A. *et al.* Short telomeres and stem cell exhaustion model Duchenne muscular dystrophy in mdx/mTR mice. *Cell* **143**, 1059–1071 (2010).
51. Crews, L. *et al.* Selective molecular alterations in the autophagy pathway in patients with Lewy body disease and in models of  $\alpha$ -synucleinopathy. *PLoS ONE* **5**, e9313 (2010).
52. Perdiguero, E. *et al.* Genetic analysis of p38 MAP kinases in myogenesis: fundamental role of p38 $\alpha$  in abrogating myoblast proliferation. *EMBO J.* **26**, 1245–1256 (2007).
53. Kametsky, L. *et al.* Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. *Bioinformatics* **27**, 1179–1180 (2011).
54. Bolte, S. & Cordelières, F. P. A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* **224**, 213–232 (2006).



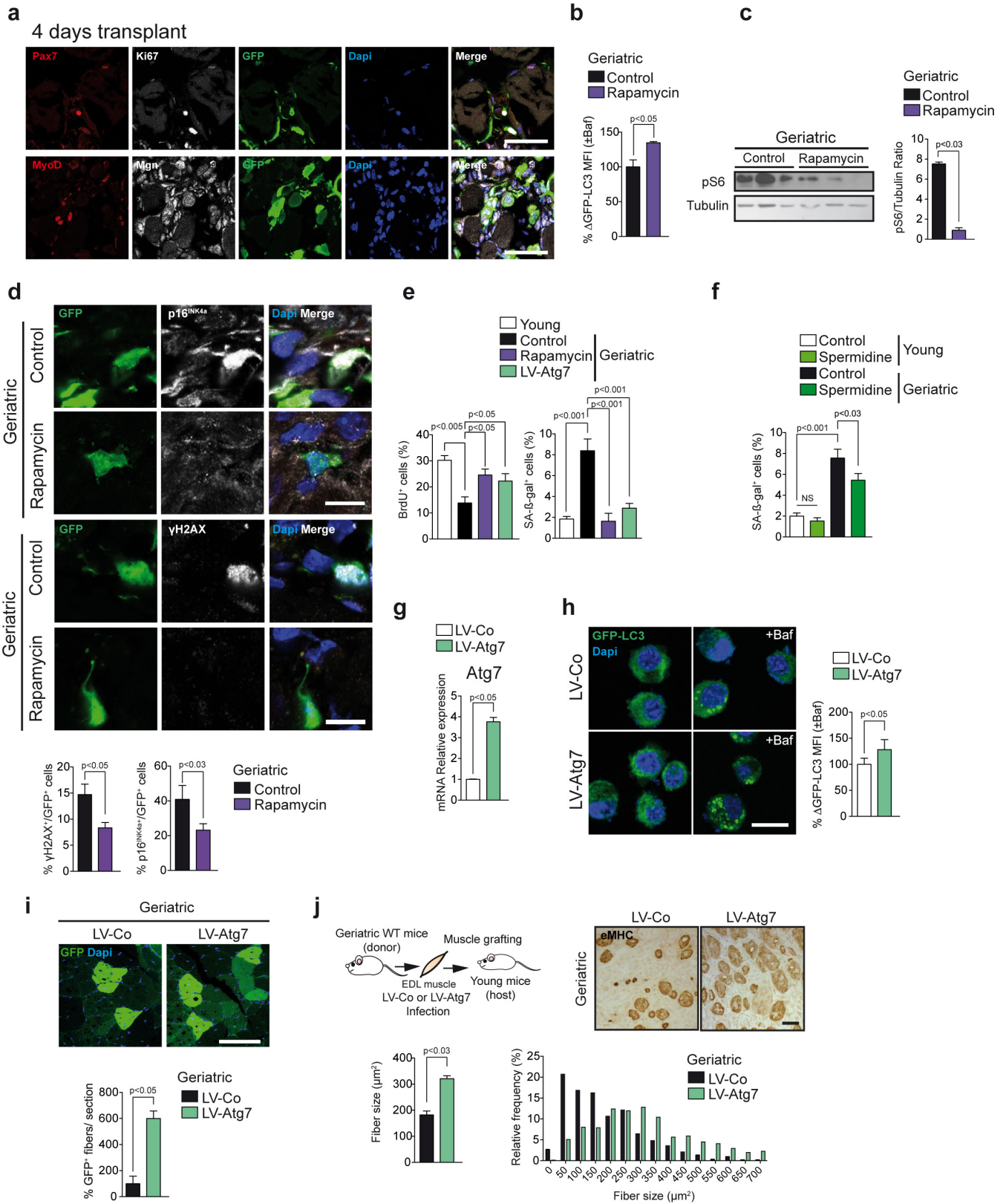
Extended Data Figure 1 | See next page for caption.



# Extended Data Figure 1 | The reduced autophagy flux in quiescent satellite cells can be increased by pharmacological treatment *in vivo*.

**a**, Venn diagrams of overlapping genes between a proteostasis gene set (See Supplementary Table 1) and genes significantly upregulated in quiescent satellite cells from the indicated publications or from our gene expression microarray data comparing freshly FACS isolated satellite cells from resting muscle, or muscles obtained 72 h after cardiotoxin (CTX) injury, from young, wild-type mice. **b**, K-means clustering analysis (performed with Gene-E, Broad Institute) of the gene expression of the autophagy-related genes during ageing. Clusters are shown with heat maps of the normalized raw data. Each column represents a different sample and each row a different gene probe. Red, increased expression; white, neutral expression; blue, decreased expression. **c**, Representative example of the FACS strategy and gating scheme for isolating satellite cells from mice in resting conditions. **d**, Pax7 and GFP immunostaining of freshly isolated satellite cells from resting muscles of young and old GFP-LC3 mice. Scale bar, 5  $\mu$ m. **e**, Electron microscopy images of young and old satellite cells on sections of resting tibialis anterior (TA) muscle of wild-type (WT) mice. Arrowheads indicate autophagic vesicles. Scale bars,

1  $\mu$ m and 0.5  $\mu$ m (right and left, respectively). **f**, Pax7 and GFP immunostaining on tissue sections from resting tibialis anterior muscles of young and old GFP-LC3 mice. Arrowheads indicate autophagic vesicles. Scale bar, 5  $\mu$ m. **g**, p62 and ubiquitin (Ub) MFI. Arrowheads, co-localization of p62 and ubiquitin aggregates. **h**, LC3 western blot of freshly isolated satellite cells from young and old, wild-type mice, treated with bafilomycin or vehicle for 4 h before collection. Graph shows LC3II quantification, after normalization with tubulin levels; for full scan see Supplementary Fig. 2. **i**, Quiescent satellite cells were freshly isolated from old, wild-type mice subjected to two weeks of rapamycin, spermidine or vehicle (control) treatment. Cells were treated (or not treated) with bafilomycin 4 h prior to analysis by immunostaining of LC3 marker. Z projections of representative fluorescence microscopy images are shown. Scale bars, 5  $\mu$ m. **j**, Representative fluorescent microscopy images from Fig. 1d. Scale bar, 5  $\mu$ m. Data show mean  $\pm$  s.e.m. Comparisons by two-sided Mann-Whitney *U*-test. *P* values are indicated. Number of samples were *n* = 3 animals per group for **a** and **b**; *n* = 35 (young) and 66 (old) cells analysed from 3 animals for **g**; *n* = 3 animals per group for **h**.

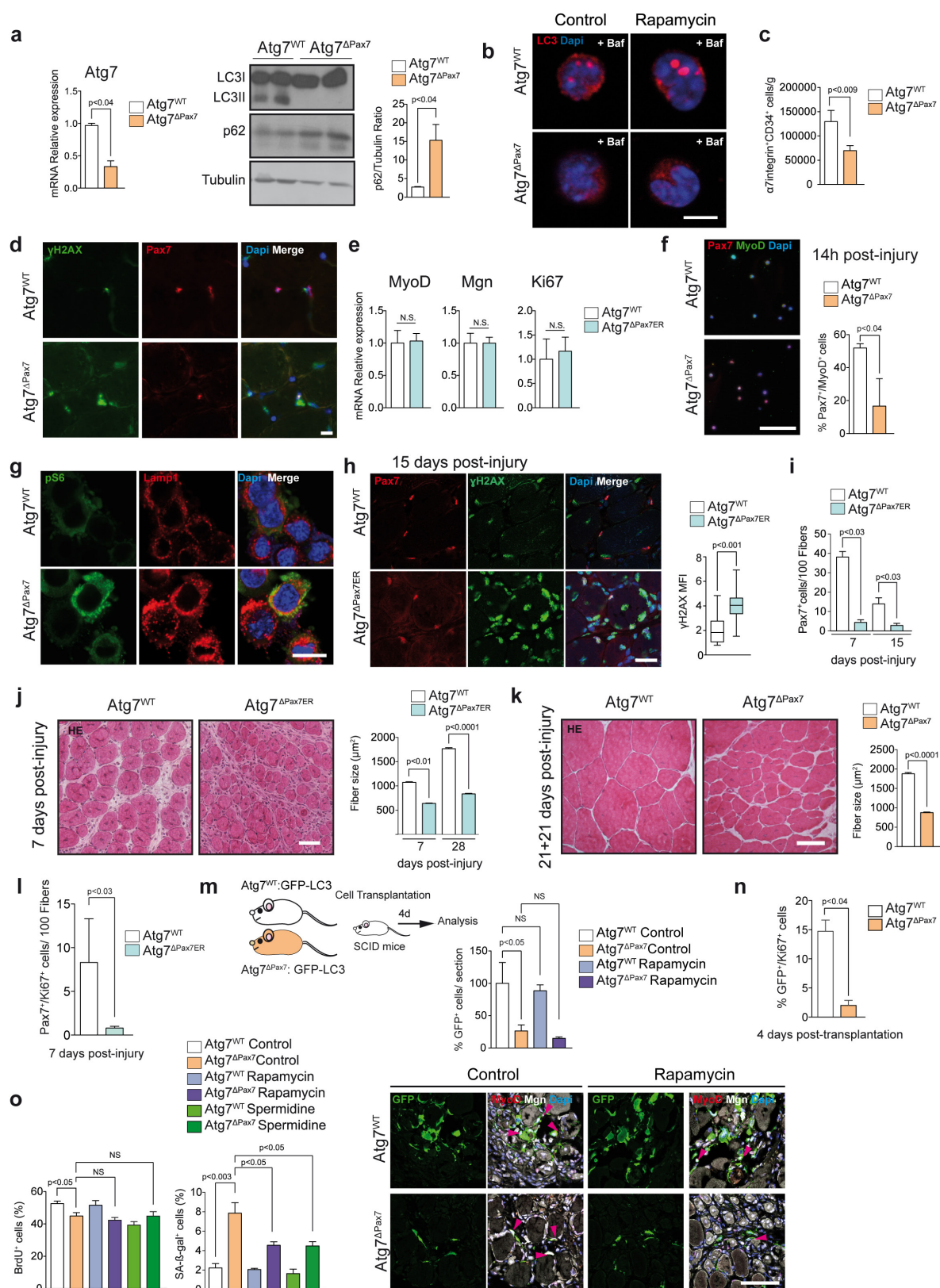


Extended Data Figure 2 | See next page for caption.



**Extended Data Figure 2 | Reinduction of autophagy rescues proliferation and reduces senescence in geriatric satellite, thus restoring regenerative capacity.** **a**, Transplanted muscles from Fig. 2c were immunostained for GFP and for Ki67, Pax7, MyoD or Mgn (to determine the distinct possible myogenic states of satellite cells in the regenerating muscle). Scale bars, 50  $\mu$ m. **b**, Autophagy flux analysed by flow cytometry in freshly isolated satellite cells from resting muscle of GFP-LC3 mice, treated for 48 h with rapamycin or vehicle (control). Satellite cells were treated with bafilomycin or vehicle for 4 h before analysis. Results are expressed as the change in GFP-LC3 MFI in bafilomycin (–) compared to bafilomycin (+) conditions. **c**, Western blot analysis of pS6 protein levels in geriatric satellite cells from wild-type mice, treated for 48 h with rapamycin or vehicle (control). Graph shows pS6 quantification, normalized to tubulin; for full scan see Supplementary Information Fig. 2. **d**, As in Fig. 2c, percentage  $\gamma$ H2AX<sup>+</sup> or p16<sup>INK4a</sup><sup>+</sup> cells from total GFP<sup>+</sup> cells were quantified. Scale bars, 10  $\mu$ m. **e**, Quantification BrdU<sup>+</sup> and senescence-associated  $\beta$ -gal<sup>+</sup> satellite cells, pre-treated as in Fig. 2c and analysed after 96 h. **f**, Quantification of senescent (senescence-associated  $\beta$ -gal<sup>+</sup>) satellite cells, isolated from young and geriatric wild-type mice, pre-treated for 48 h with spermidine or vehicle (control) and cultured for 96 h. **g**, Quantitative real-time PCR (RT-qPCR) analysis of Atg7 expression on satellite cells infected with LV-Atg7 or LV-control

(LV-Co), and cultured for 96 h. **h**, GFP-LC3 satellite cells were infected with LV-Atg7 or LV-Co and treated with bafilomycin or vehicle for 4 h before analysis. Autophagy flux was analysed by flow cytometry and represented as in **b**. Representative images are shown. Scale bar, 10  $\mu$ m. **i**, Muscle regeneration experiment by satellite cell transplantation. An equal number of satellite cells from young and geriatric mice infected with a lentivirus overexpressing the *Atg7* gene (LV-Atg7) or a lentivirus control (LV-Co), which also expressed GFP, were transplanted into injured muscle of young immunodeficient mice, and collected 28 days later. GFP expression in muscles was analysed by immunostaining. Quantification of GFP<sup>+</sup> cells (fibres) per muscle field versus transplanted control-treated satellite cells. Representative images are shown. Scale bar, 75  $\mu$ m. **j**, EDL geriatric muscles, infected with LV-Atg7 or LV-Co, and grafted on recipient mouse muscle. Regeneration was analysed 8 days later. Frequency distribution of regenerating fibres by size. Scale bar, 25  $\mu$ m. Data are mean  $\pm$  s.e.m. Comparisons by two-sided Mann–Whitney *U*-test. *P* values are indicated. Number of samples were *n* = 60,000 cells analysed from 3 animals for **b**; *n* = 3 animals per group for **c**; *n* = 5 engraftments per group for **d**; *n* = 3 animals per group for **e–g**; *n* = 60,000 cells analysed from 3 animals for **h**; *n* = 3 engraftments per group for **i**; *n* = 4 engraftments per group for **j**.



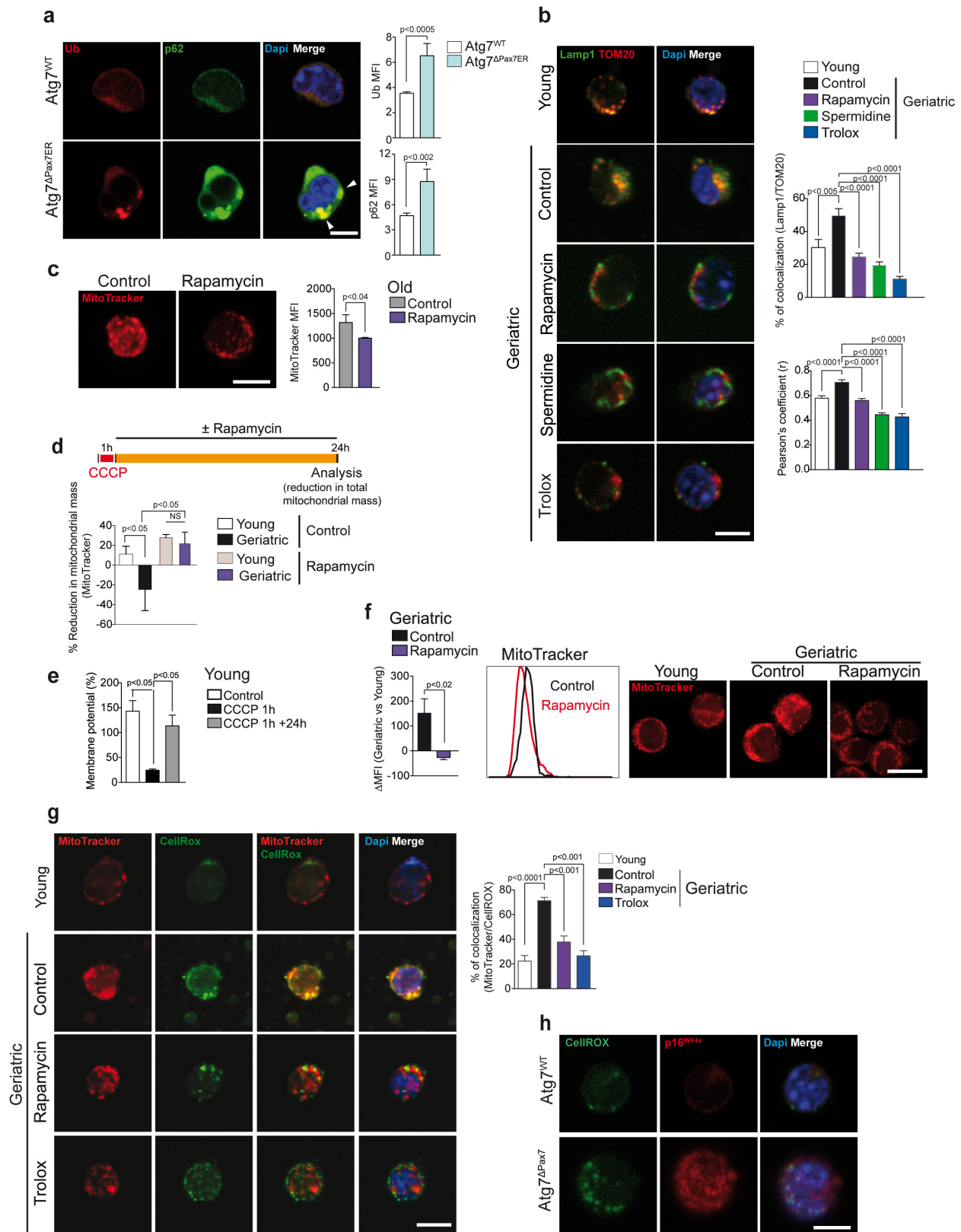
Extended Data Figure 3 | See next page for caption.



### Extended Data Figure 3 | Genetic impairment of autophagy in young quiescent satellite cells leads to premature senescence and impaired muscle regeneration.

**a**, RT-qPCR analysis of *Atg7* expression and western blot analysis of LC3, p62 and tubulin of satellite cells isolated from *Atg7*<sup>WT</sup> and *Atg7*<sup>ΔPax7</sup> mice. Graph shows the quantification of p62 normalized to tubulin; for full scan see Supplementary Fig. 2. **b**, Quiescent satellite cells were freshly isolated from *Atg7*<sup>WT</sup> and *Atg7*<sup>ΔPax7</sup> mice which had been subjected to two weeks of rapamycin or vehicle (control) treatment *in vivo*. Cells were treated (or controls were untreated) with bafilomycin 4 h before analysis by fluorescence microscopy. Z projections of representative fluorescence microscopy images are shown. Scale bar, 5 μm. **c**, Quantification of satellite cells in resting muscle of three-month-old *Atg7*<sup>WT</sup> and *Atg7*<sup>ΔPax7</sup> mice by flow cytometry analysis (α7 integrin<sup>+</sup>CD34<sup>+</sup> cells per gram of muscle tissue). **d**, Representative fluorescent microscopy images from Fig. 3d. Scale bar, 10 μm. **e**, RT-qPCR analysis of MyoD, Mgn and Ki67 expression in freshly isolated quiescent satellite cells from resting muscle of *Atg7*<sup>WT</sup> and *Atg7*<sup>ΔPax7ER</sup> mice, 7 days after tamoxifen treatment. **f**, Percentage of activated satellite cells (Pax7<sup>+</sup>/MyoD<sup>+</sup>) from the total Pax7<sup>+</sup> cells (FACS-isolated 14-h post-injury from **a**). Scale bar, 50 μm. **g**, pS6 and Lamp1 immunostaining of cells from **a**. Scale bar, 10 μm. **h**, γH2AX protein levels per nucleus in Pax7<sup>+</sup> satellite cells in tibialis anterior muscles of *Atg7*<sup>WT</sup> and *Atg7*<sup>ΔPax7ER</sup> mice, 15 days post-injury. Representative images are shown. Scale bar, 25 μm. **i**, Pax7<sup>+</sup> satellite cells were quantified following immunostaining on regenerating muscles of *Atg7*<sup>WT</sup> and *Atg7*<sup>ΔPax7ER</sup>

mice 7 days and 15 days after cardiotoxin injury. **j**, Representative images of haematoxylin and eosin staining of muscles at 7 days post-injury on muscles of *Atg7*<sup>WT</sup> and *Atg7*<sup>ΔPax7ER</sup> mice. Fibre size of central-nucleated myofibres at 7 days and 28 days post-injury is quantified. Scale bar, 50 μm. **k**, Tibialis anterior muscles of *Atg7*<sup>WT</sup> and *Atg7*<sup>ΔPax7</sup> mice were injured by cardiotoxin injection and 21 days later these muscles were reinjured and then subsequently analysed 21 days later (21 + 21 days post-injury). The size of central-nucleated myofibres was quantified. Representative images are shown. Scale bar, 50 μm. **l**, Pax7<sup>+</sup> and Ki67<sup>+</sup> double-positive satellite cells were quantified following immunostaining on regenerating muscles of *Atg7*<sup>WT</sup> and *Atg7*<sup>ΔPax7ER</sup> mice 7 days after cardiotoxin injury. **m**, An equal number of quiescent satellite cells from *Atg7*<sup>WT</sup>:GFP-LC3 and *Atg7*<sup>ΔPax7</sup>:GFP-LC3 mice (two weeks ± rapamycin pre-treatment), transplanted as in Fig. 2c, and immunostained with the indicated antibodies 4 days later. Quantification of GFP<sup>+</sup> cells per muscle field. Values relative to transplanted young cells (100%). Representative images are shown. Scale bar, 75 μm. **n**, Percentage of GFP<sup>+</sup> cells that are also Ki67<sup>+</sup> cells in muscles from **m**. **o**, Quantification of proliferating (BrdU<sup>+</sup>) and senescent (SA-β-gal<sup>+</sup>) satellite cells, isolated from *Atg7*<sup>WT</sup> and *Atg7*<sup>ΔPax7</sup>, pre-treated for 48 h with spermidine or rapamycin (or control vehicle) and cultured for 96 h. Data show mean ± s.e.m. Comparisons by two-sided Mann-Whitney *U*-test. *P* values are indicated. The number of samples were *n* = 3 animals per group (**a**); *n* = 7 animals per group (**c**); *n* = 3 animals per group (**e-l**); *n* = 4 engraftments per group (**m, n**); *n* = 3 animals per group (**o**).

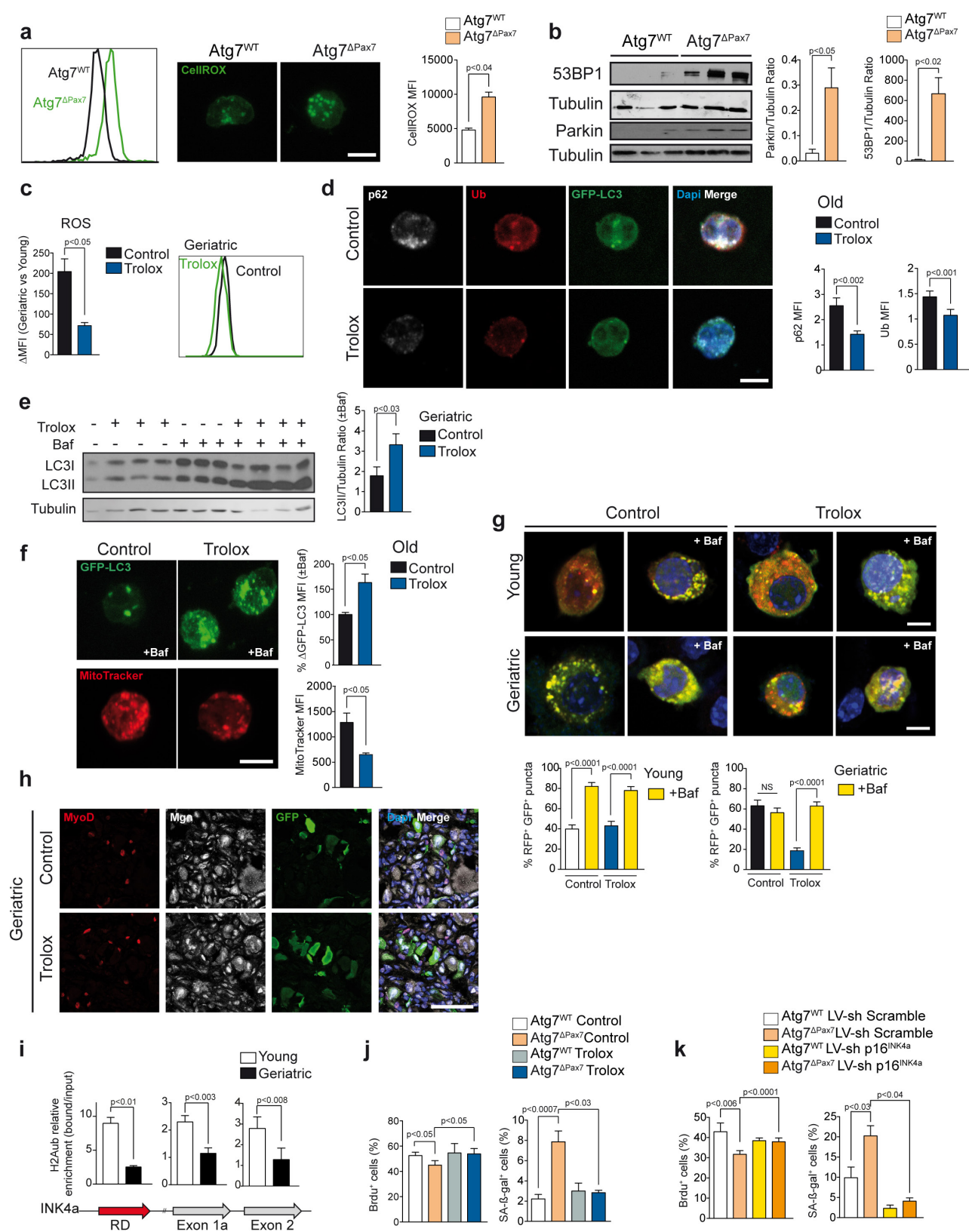


Extended Data Figure 4 | See next page for caption.



**Extended Data Figure 4 | Autophagy loss in satellite cells causes dysfunctional mitophagy and mitochondria accumulation, leading to increased ROS and senescence.** **a**, p62 and ubiquitin immunostaining on freshly isolated satellite cells from resting muscle of three-month-old  $Atg7^{WT}$  and  $Atg7^{\Delta Pax7^{ER}}$  mice, one month after tamoxifen treatment. Arrowheads indicate co-localization of p62 and Ub aggregates. Representative images are shown. Scale bar, 5  $\mu$ m. **b**, TOM20 and Lamp1 immunostaining of quiescent satellite cells isolated from young and geriatric WT mice. Mice were subjected to two weeks of rapamycin, spermidine or Trolox (or vehicle) treatment before analysis. Co-localization was calculated as the area occupied by the immunofluorescence co-localizing staining on images with respect to the total cellular area. The Pearson's coefficient ( $r$ ) was used to analyse the correlation of the intensity values of green and red pixels in the dual-channel images. The  $z$  projections of representative fluorescence microscopy images are shown. Scale bar, 5  $\mu$ m. **c**, Mitochondria quantification by MitoTracker in quiescent satellite cells of old mice, treated with rapamycin or vehicle for two weeks. **d**, Mitochondria (MitoTracker labelling) in young or geriatric cells. Satellite cells, were pre-treated with CCCP for 1 h (see Methods) and  $\pm$  rapamycin for 24 h. Percentage of MitoTracker MFI reduction  $\pm$  rapamycin. **e**, For the mitochondrial membrane potential analysis, satellite cells were freshly

isolated from young wild-type mice and treated for 1 h with CCCP or DMSO (control). Membrane potential (TMRM MFI/MitoTracker Green MFI ratio) of cells was calculated by flow cytometry analysis at 1 h and 24 h after CCCP treatment (being 100% the membrane potential value of control satellite cells). **f**, Mitochondria content was quantified by MitoTracker staining of satellite cells from young and geriatric wild-type mice treated with rapamycin or vehicle (control) for 48 h. The  $z$  projections of representative fluorescence microscopy images are shown. Scale bar, 5  $\mu$ m. **g**, Mitochondria and ROS detection by MitoTracker and CellROX staining, respectively. Co-localization was calculated as in **b**. The  $z$  projections of representative fluorescence microscopy images are shown. Scale bar, 5  $\mu$ m. **h**, Representative images of freshly isolated satellite cells from resting muscle of three-month-old  $Atg7^{WT}$  and  $Atg7^{\Delta Pax7}$  mice stained with CellROX fluorescent dye and p16<sup>INK4a</sup> antibody. Scale bar, 5  $\mu$ m. Data are mean  $\pm$  s.e.m. Comparisons by two-sided Mann–Whitney  $U$ -test.  $P$  values are indicated. Number of samples were  $n = 36$  ( $Atg7^{WT}$ ) and  $n = 38$  ( $Atg7^{\Delta Pax7^{ER}}$ ) cells analysed from 3 animals (**a**);  $n = 23$  (young),  $n = 24$  (control),  $n = 42$  (rapamycin);  $n = 28$  (spermidine) and  $n = 21$  (Trolox) cells analysed from 3 animals (**b**);  $n = 60,000$  cells analysed from 3 animals (**c**);  $n = 40,000$  cells analysed from 4 animals (**d**);  $n = 30,000$  cells analysed from 3 animals (**e**, **f**);  $n = 18$  (young), 21 (control), 15 (rapamycin) and 13 (Trolox) cells analysed from 3 animals (**g**).



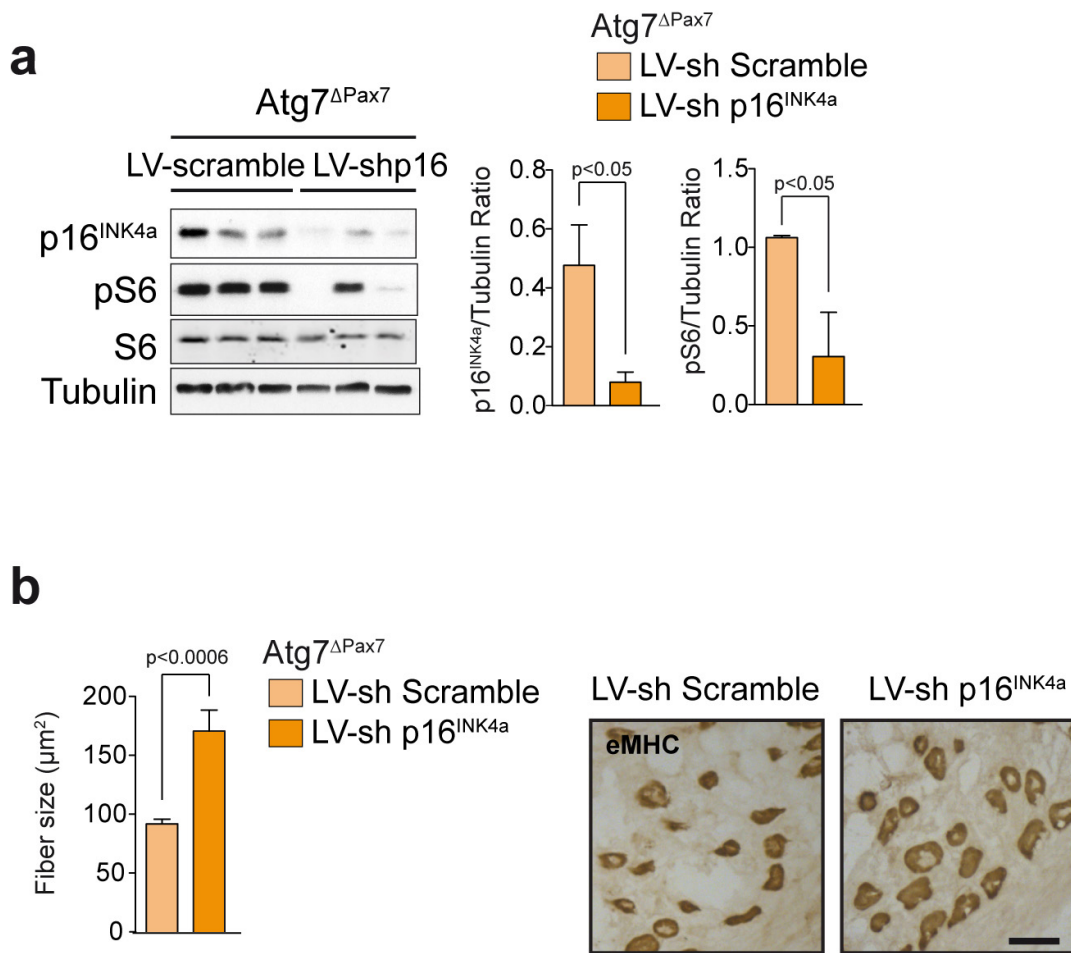
Extended Data Figure 5 | See next page for caption.



### Extended Data Figure 5 | ROS inhibition in autophagy-impaired aged and *Atg7* null satellite cells significantly restores cell proteostasis.

**a**, ROS-level quantification in quiescent satellite cells from three-month-old *Atg7*<sup>WT</sup> and *Atg7*<sup>ΔPax7</sup> mice by CellROX flow cytometry. Representative images are shown. Scale bar, 5 μm. **b**, Western blot analysis of 53BP1 and parkin in satellite cells isolated from three-month-old *Atg7*<sup>WT</sup> and *Atg7*<sup>ΔPax7</sup> mice. Tubulin control is the same tubulin control for Fig. 3g. Graph shows quantification of 53BP1 and parkin protein normalized to tubulin; for full scan see Supplementary Information Fig. 1. **c**, Quantification of ROS levels for satellite cells isolated from young and geriatric WT mice by flow cytometry using CellROX fluorescent dye. Satellite cells were treated with Trolox or vehicle (control) for 48 h before analysis. Results are represented as variation of MFI between young and geriatric satellite cells. **d**, Quantification of p62 and ubiquitin protein levels on immunostained freshly isolated satellite cells from resting muscle of old wild-type mice, *in vivo* treated for 2 weeks with Trolox or vehicle (control). Representative images are shown. Scale bar, 5 μm. **e**, Western blot analysis of LC3 and tubulin in satellite cells isolated from geriatric WT mice and treated for 48 h with Trolox or vehicle (control), in the absence or presence of bafilomycin for 4 h before analysis. Graph shows quantification of LC3II protein normalized to tubulin; for full gel scan see Supplementary Information Fig. 2. **f**, Autophagy flux and mitochondria in satellite cells from GFP-LC3 mice (two weeks with or without Trolox treatment). Satellite cells treated for 4 h ± bafilomycin treatment. Representative images are shown. Scale bar, 5 μm. **g**, The mRFP-GFP-LC3 plasmid was transfected into young or geriatric satellite

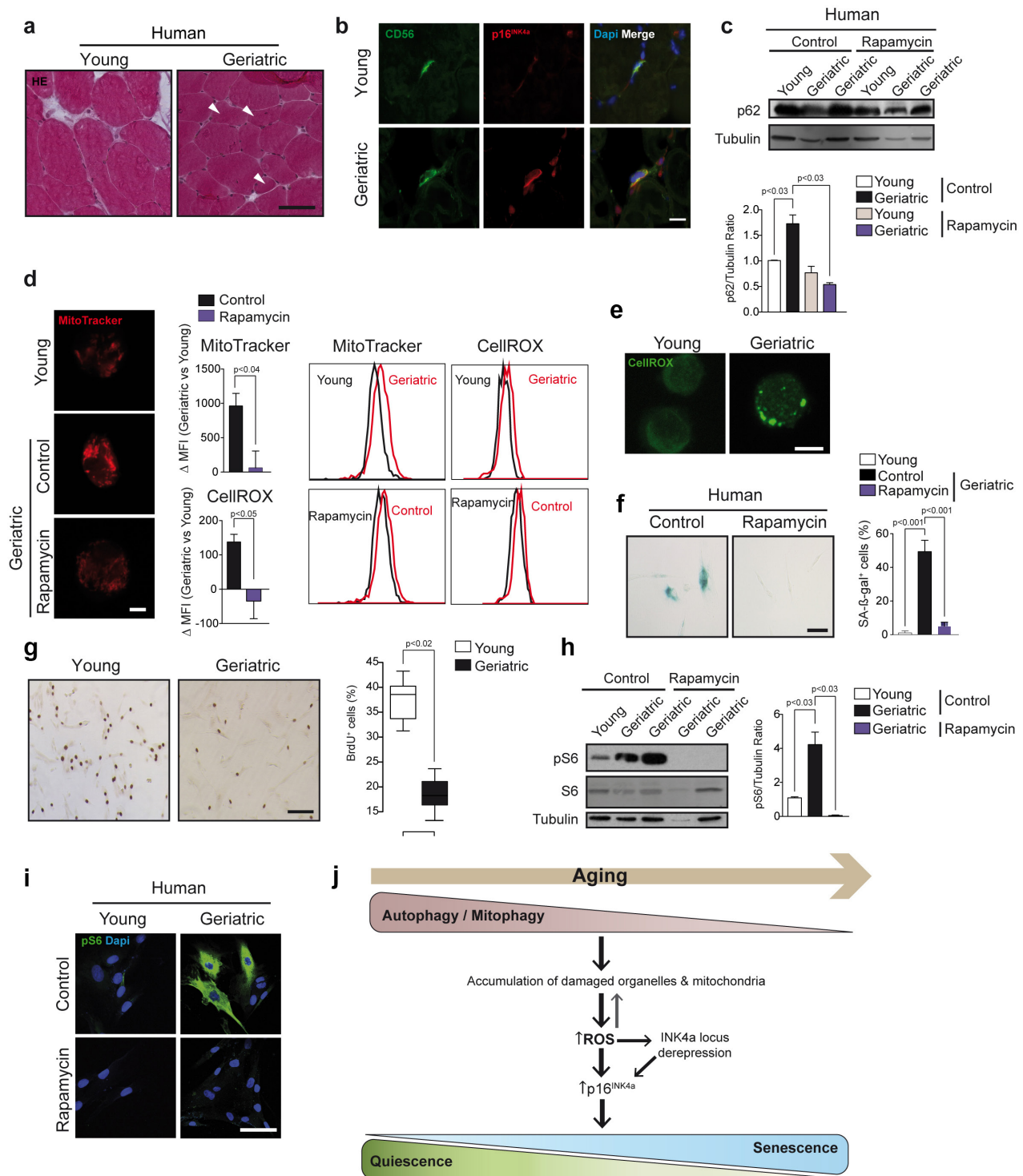
cells, with 48 h treatment ± Trolox and then 4 h treatment ± bafilomycin, prior to fixation. The percentage of autophagosomes was quantified as in Fig. 2a. **h**, Muscle regeneration using geriatric satellite cell transplantation. An equal number of freshly isolated geriatric satellite cells, infected with GFP lentivirus and treated for 48 h with Trolox or vehicle, were transplanted into injured muscle of young immunodeficient mice. Four days later, muscles were collected and immunostained for GFP, MyoD and Mgn (to determine the possible myogenic states of satellite cells in the regenerating muscle). Representative images are shown. Scale bar, 50 μm. **i**, ChIP analysis for H2AK119ub (H2Aub) in satellite cells isolated from young and geriatric wild-type mice. **j**, Quantification of proliferating (BrdU<sup>+</sup>) and senescent (SA-β-gal<sup>+</sup>) satellite cells isolated from *Atg7*<sup>WT</sup> and *Atg7*<sup>ΔPax7</sup> mice treated 48 h with Trolox or vehicle (control) and cultured for 96 h. **k**, Quantification of proliferating (BrdU<sup>+</sup>) and senescent (senescence-associated β-gal<sup>+</sup>) satellite cells isolated from *Atg7*<sup>WT</sup> and *Atg7*<sup>ΔPax7</sup> mice and infected with LV-sh p16<sup>INK4a</sup> or LV-sh scramble, and cultured for 96 h. Data show mean ± s.e.m. Comparisons by two-sided Mann–Whitney *U*-tests. *P* values are indicated. Number of samples were *n* = 60,000 cells analysed from 3 animals (**a**); *n* = 3 animals per group (**b**); *n* = 60,000 cells analysed from 3 animals (**c**); *n* = 36 (control) and *n* = 35 (Trolox) cells analysed from 3 animals (**d**); *n* = 3 animals per group (**e**); *n* = 60,000 cells analysed from 3 animals (**f**); *n* = 21 (young), *n* = 20 (young, Trolox), *n* = 19 (young, + bafilomycin), *n* = 18 (young, Trolox + bafilomycin), *n* = 21 (geriatric), *n* = 19 (geriatric, Trolox), *n* = 15 (geriatric, + bafilomycin) and *n* = 37 (geriatric, Trolox + bafilomycin) cells analysed from 3 animals (**g**); *n* = 3 animals per group (**i–k**).



**Extended Data Figure 6 | Effects of p16<sup>INK4a</sup> silencing in autophagy-impaired young murine satellite cells.** **a**, Western blotting quantification of Atg7<sup>ΔPax7</sup> satellite cells, infected with lentiviral LV-sh-p16<sup>INK4a</sup> or LV-sh-scramble and analysed 96 h later; for full gel scan see Supplementary Fig. 2. **b**, Atg7<sup>WT</sup> and Atg7<sup>ΔPax7</sup> EDL, infected with LV-sh-p16<sup>INK4a</sup> or

LV-sh-scramble, and grafted as Extended Data Fig. 2j. Representative eMHC-immunostaining. Scale bar, 25 μm. Data show mean ± s.e.m. Comparisons by two-sided Mann-Whitney *U*-test. *P* values are indicated. The number of samples were *n* = 3 animals per group (**a**) and *n* = 4 engraftments per group (**b**).





**Extended Data Figure 7 | Impaired autophagic flux in human geriatric satellite cells.** **a**, Representative images of haematoxylin and eosin staining of human muscle biopsies from young (25 years old) and geriatric (95 years old) donors in resting conditions. Arrowheads indicate atrophic myofibres. Scale bar, 50  $\mu$ m. **b**, CD56 and p16<sup>INK4a</sup> immunostaining on human muscle sections of samples described in **a**. Scale bar, 10  $\mu$ m. **c**, Western blotting analysis of p62 protein in human satellite cells from young (about 25 years old) and geriatric (over 75 years old) donors, treated for 48 h  $\pm$  rapamycin; for full gel scan see Supplementary Fig. 2. **d**, ROS and mitochondrial content analysis in human cells from treated for 48 h  $\pm$  rapamycin. Graphs show MFI variation. Scale bar, 5  $\mu$ m. **e**, Representative images from CellROX staining from **d**. Scale bar, 5  $\mu$ m. **f**, Quantification of SA- $\beta$ -gal<sup>+</sup> human cells treated for 48 h  $\pm$  rapamycin. Quantification was carried out 96 h after treatment. Scale bar, 200  $\mu$ m.

**g**, Quantification of proliferating (BrdU<sup>+</sup>) young and geriatric human satellite cells in culture. Representative pictures are shown. Scale bar, 25  $\mu$ m. **h**, Western blot analysis of pS6, total S6 and tubulin in young and geriatric human satellite cells treated for 48 h with rapamycin or vehicle (control). Graphs show p62 quantification normalized to tubulin; for full scan see Supplementary Information Fig. 2. **i**, Immunostaining of pS6 in young and geriatric human satellite cells treated as in **h**. Scale bar, 75  $\mu$ m. **j**, Scheme showing the proposed model of how age-impaired autophagy leads to muscle stem-cell senescence and regenerative decline. Data show mean  $\pm$  s.e.m. Comparisons by two-sided Mann-Whitney *U*-tests. *P* values are indicated. The number of samples were  $n = 3$  human donors per group (**a**–**c**);  $n = 60,000$  cells analysed from 3 human donors per group (**f**–**h**).

# Substantial contribution of extrinsic risk factors to cancer development

Song Wu<sup>1,2</sup>, Scott Powers<sup>1,2,3</sup>, Wei Zhu<sup>1,2</sup> & Yusuf A. Hannun<sup>2,3,4,5</sup>

**Recent research has highlighted a strong correlation between tissue-specific cancer risk and the lifetime number of tissue-specific stem-cell divisions. Whether such correlation implies a high unavoidable intrinsic cancer risk has become a key public health debate with the dissemination of the ‘bad luck’ hypothesis. Here we provide evidence that intrinsic risk factors contribute only modestly (less than ~10–30% of lifetime risk) to cancer development. First, we demonstrate that the correlation between stem-cell division and cancer risk does not distinguish between the effects of intrinsic and extrinsic factors. We then show that intrinsic risk is better estimated by the lower bound risk controlling for total stem-cell divisions. Finally, we show that the rates of endogenous mutation accumulation by intrinsic processes are not sufficient to account for the observed cancer risks. Collectively, we conclude that cancer risk is heavily influenced by extrinsic factors. These results are important for strategizing cancer prevention, research and public health.**

Cancers were once thought to originate from mature tissue cells that underwent dedifferentiation in response to cancer progression<sup>1</sup>. Today, cancers are proposed to originate from the malignant transformation of normal tissue progenitor and stem cells<sup>2,3</sup>, although this is not wholly accepted<sup>4</sup>. Nevertheless, recent research has highlighted a strong correlation of 0.81 between tissue-specific cancer risk and the lifetime population size in cumulative number of cell divisions of tissue-specific stem cells<sup>5</sup>. However, there has been controversy regarding the conclusion that this correlation implies a very high unavoidable risk for many cancers that is due solely to the intrinsic baseline population size of tissue-specific stem cells<sup>6,7</sup>. Many arguments against the ‘bad luck’ hypothesis have been made<sup>5–13</sup>, yet none of these have offered specific alternatives to quantitatively evaluate the contribution of extrinsic risk factors in cancer development. Applying several distinct modelling approaches, here we provide strong evidence that unavoidable intrinsic risk factors contribute only modestly (less than ~10–30%) to the development of many common cancers.

We made the conservative and yet conventional assumption that errors occurring during the division of cells, being routes of malignant transformation, can be influenced by both intrinsic processes as well as extrinsic factors (Fig. 1). ‘Intrinsic processes’ include those that result in mutations due to random errors in DNA replication, whereas ‘extrinsic factors’ are environmental factors that affect mutagenesis rates (such as ultraviolet (UV) radiation, ionizing radiation and carcinogens). For example, radiation can cause DNA damage, which would primarily result in deleterious mutations with functional consequences on cancer development only after cell division. Therefore, extrinsic factors may act through the accumulation of genetic alterations during cell division to increase cancer risk. Accordingly, cancer risk would result from those apparently uncontrollable intrinsic processes (Fig. 1, arrow 1) as well as from those highly modifiable and thus preventable extrinsic factors (Fig. 1, arrow 2).

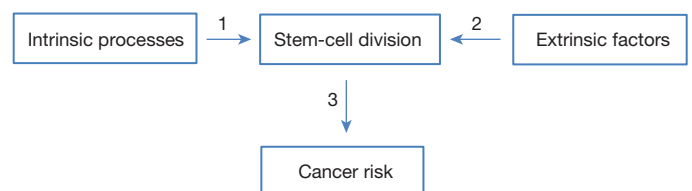
## Correlation cannot differentiate risks

According to the above hypothesis, both intrinsic and extrinsic factors can impart cancer risk through the accumulation of these errors, especially the ‘driver mutations’ (Fig. 1, arrow 3). As such, a correlational

analysis between cancer risk and cell division, for either stem or non-stem cells, is unable to differentiate between the contributions of intrinsic and extrinsic factors. This is best illustrated through a thought experiment where we consider a hypothetical scenario of a sudden global emergence of a very potent mutagen, such as a strong radiation burst from a nuclear fallout, which quadruples the lifetime risks for all cancers. In this scenario, it transpires that the proportion of cancer risk caused by intrinsic random errors would be small (at most one-quarter if we assume all of the original risk was due to intrinsic processes). However, if we conduct regression analyses on either the new hypothetical cancer risks or the current cancer risks as reported, against the number of stem-cell divisions<sup>5</sup>, the correlations from both cases would be 0.81 (Fig. 2). This thought experiment negates the ability of the correlation to detect solely the contribution of intrinsic factors as it cannot distinguish between intrinsic and extrinsic factors. Thus, it argues against the implication that around two-thirds of variation could be explained by division-related random intrinsic errors.

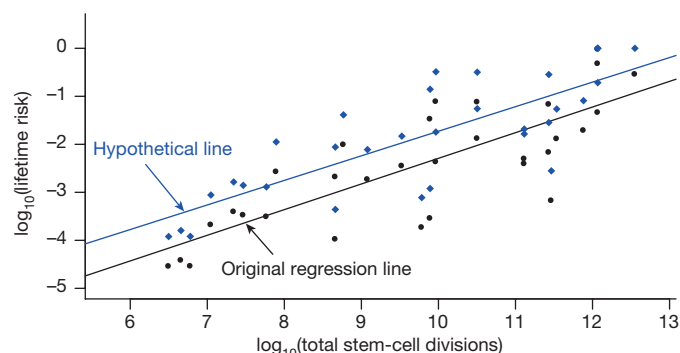
## Lower bound intrinsic risk line

The above conclusion then raises the question of what proportion of total cancer risk is due to extrinsic versus intrinsic factors. In a data-driven approach, we first re-examined the quantitative relationship between the observed lifetime cancer risk and the divisions of the



**Figure 1 | Schematic showing how intrinsic processes and extrinsic factors relate to cancer risks through stem-cell division.** This hypothesis maintains the strong role of stem-cell division in imparting cancer risk, but it also illustrates the potential contributions of both intrinsic and extrinsic factors operating through stem-cell division. Other effects, for example, through division of non-stem cells, are considered later in this analysis.

<sup>1</sup>Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, New York 11794, USA. <sup>2</sup>Stony Brook Cancer Center, Stony Brook University, Health Sciences Center, Stony Brook, New York 11794, USA. <sup>3</sup>Department of Pathology, Stony Brook University, Health Sciences Center, Stony Brook, New York 11794, USA. <sup>4</sup>Department of Medicine, Stony Brook University, Health Sciences Center, Stony Brook, New York 11794, USA. <sup>5</sup>Department of Biochemistry, Stony Brook University, Health Sciences Center, Stony Brook, New York 11794, USA.



**Figure 2 | Correlation analysis of stem-cell division and cancer risk does not distinguish contribution of extrinsic versus intrinsic factors to cancer risk.** The black dots are data from figure 1 (also shown in supplementary table 1) of Tomasetti & Vogelstein<sup>5</sup>, and the black line shows their original regression line. The blue diamonds represent the hypothesized quadrupled cancer risks due to hypothetical exposure to an extrinsic factor such as radiation. The blue regression line for the hypothetical risk data maintains the same correlation as the original black line, albeit reflecting a much higher contribution of extrinsic factors to cancer risk.

normal tissue stem cells as reported<sup>5</sup>, with a distinct alternative method. Our rationale was that intrinsic risk, or indeed its upper bound, can be better estimated by the lowest boundary on the plots of cancer risk versus total tissue stem-cell divisions (Fig. 3a, red 'intrinsic' risk line), meaning that intrinsic cancer risk should be determined by the cancer incidence for those cancers with the least risk in the entire group controlling for total stem-cell divisions (Fig. 3a, red dots). The argument here is that cancers with the same number of stem-cell divisions should share the same base of intrinsic cancer risk (if the relationship is causal); if one or more cancers would feature a much higher cancer incidence, for example, lung cancer among smokers versus non-smokers, then this probably reflects additional (and probably extrinsic) risk factors (smoking in this case). One could argue that the low-incidence tumour types may have lower incidences because of additional genetic repair mechanisms that restrict evolving malignant cells from accumulating sufficient numbers of genetic alterations required to become fully tumorigenic; however, without more specific data on the operation of repair mechanisms, these could drive the risk up or down, depending on whether they are less or more efficient in any particular tissue. According to our hypothesis, intrinsic risk from stem-cell divisions would define the lowest bound for a given number of stem-cell divisions, therefore we define an 'intrinsic' risk line for stem-cell divisions by regressing the smallest cancer risks on any given number of stem-cell divisions (Fig. 3a, red line). The 'intrinsic' risk lines themselves are still probably overestimates for the intrinsic risk; however, we should suspect that any cancer risk above that line implies additional biologic determinants, on the basis of which we can compute the percentage of cancer risk not explained by intrinsic 'randomness'. As shown in Fig. 3a, most cancer types have very high excess risks relative to the 'intrinsic' risk line, indicating large proportions of risks that are unaccounted for by the intrinsic factors, typically larger than 90%. Moreover, these estimated excess risks are very robust: with plausible measurement errors added to the total stem-cell divisions, the resulting excess risks remain essentially intact (Extended Data Table 1).

### Extrinsic risks by tissue cell turnover

Although we performed the initial analysis from a 'stem-cell theory' point of view, we wanted to evaluate if our results are dependent on this specific theory or independent of it. Furthermore, the lack of reliable data on human tissue stem-cell dynamics is a notable concern (see Supplementary Information), rendering the analysis in Fig. 3a less determinate. Thus, we separately collected data for the total number of tissue cell divisions that is based on homeostatic tissue cell numbers

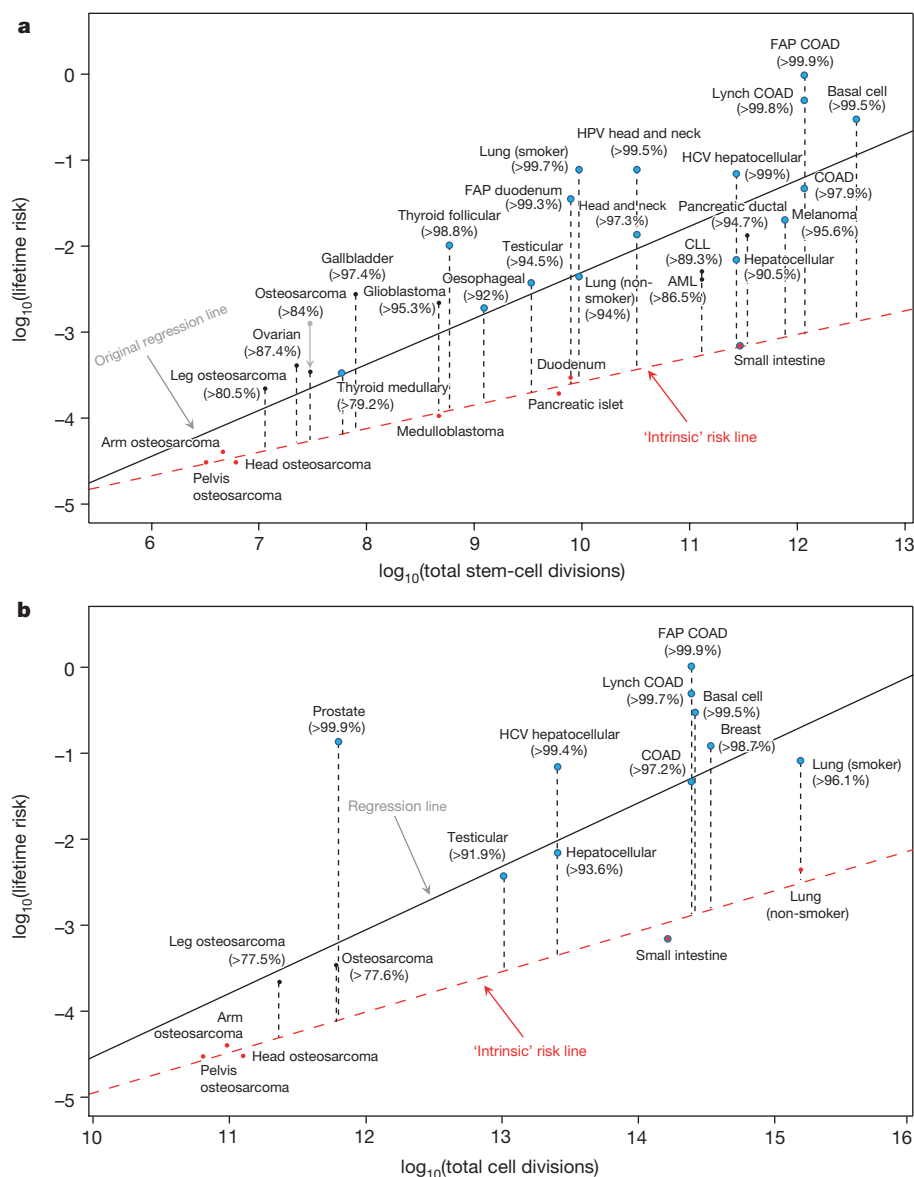
and their turnover rates (see Supplementary Information), and analysed the relationship of cancer risk versus total tissue cell divisions (Fig. 3b). This approach allows for every dividing cell to be a potential cancer-initiating cell, which would be an application of another cell-of-origin theory of cancer whereby tumours may originate from a hierarchy of cells, from stem cells to committed progenitor cells to differentiated cells<sup>4</sup>. Mathematically, this can also be considered as an extreme form of stem-cell theory where the fraction of stem cells is 1 (this latter formulation then provides an upper bound of the effects of the size of the stem-cell population on cancer risk and the role of extrinsic factors). The regression analysis between cancer risk and total tissue cell division shows a high correlation of 0.75, establishing a strong quantitative relationship between cancer risk and total cell division. To dissect the extrinsic versus intrinsic risks, we applied the same rationale and regressed the smallest cancer risks on any given number of cell divisions (Fig. 3b, red line). Although we could only find reliable turnover data for a subset of tissues, it is remarkable that the conclusion drawn here is nearly identical to that in Fig. 3a; that is, large proportions of risks that may not be attributable to intrinsic factors are mostly higher than 90%. It is important to note that here we included breast and prostate cancers—two high-incidence cancers missing in the original stem-cell analysis<sup>5</sup>. Again, plausible measurement errors have been added to the total cell divisions, and the excess risks remained almost identical (Extended Data Table 1). In summary, irrespective of whether a subpopulation or all dividing cells contribute to cancer, these results indicate that intrinsic factors do not play a major causal role.

### Epidemiological evidence

In parallel, numerous epidemiological studies have established strong evidence that many cancers have substantial risk proportions attributed to environmental exposures (Extended Data Table 2). Particularly, for breast and prostate cancers, it has long been observed that large international geographical variations exist in their incidence rates (for example, Western Europe has the highest incidence of breast cancer, which is almost 5 times higher than areas such as Eastern Asia or Middle Africa; Australia/New Zealand has the highest incidence of prostate cancer, which is almost 25 times higher than areas such as South-Central Asia)<sup>14</sup>, and immigrants moving from countries with lower cancer incidence to countries with higher cancer rates soon acquire the higher risk of their new country<sup>15,16</sup>. While several risk factors have been identified for these cancers, no single one can account for their substantial extrinsic risk proportions, suggesting complex mechanisms for their aetiologies. Colorectal cancer is a high-incidence cancer that is widely considered to be an environmental disease<sup>17</sup>, with an estimated 75% or more of colorectal cancer risk attributable to diet<sup>18</sup>. For many other cancers, known environmental risk factors have also been identified. For example, for melanoma the risk ascribed to sun exposure is around 65–86%<sup>19</sup>, and for non-melanoma basal and squamous skin cancers ~90% is attributable to UV radiation<sup>20</sup>. At least 75% of oesophageal cancer, or head and neck cancer, is caused by tobacco and alcohol<sup>21,22</sup>. It is also well known that certain pathogens may markedly increase the risk of cancers. For instance, human papilloma virus may cause ~90% of cervical cancer cases<sup>23</sup>, ~90% of anal cancer cases<sup>24</sup> and ~70% of oropharyngeal cancer cases<sup>25</sup>; hepatitis B and C may account for ~80% of hepatocellular carcinoma cases<sup>26</sup>; and *Helicobacter pylori* may be responsible for 65–80% of gastric cancer cases<sup>27</sup>. These, along with many other reports, provide direct evidence that environmental factors play important roles in cancer incidence and they are modifiable through lifestyle changes and/or vaccinations.

Additionally, analyses of data from the Surveillance, Epidemiology, and End Results Program (SEER) in the USA between 1973–2012 demonstrate that while many cancers have declining or maintain relatively consistent age-adjusted incidence rates (for example, cervical, gallbladder and oesophageal cancers, Extended Data Fig. 1), incidences of some cancers (including melanoma, thyroid, kidney, liver, thymus,





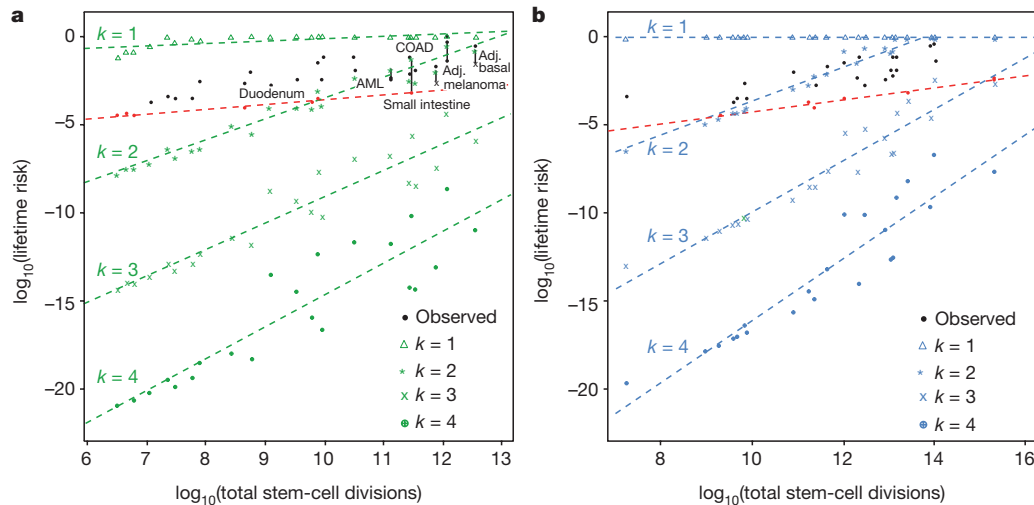
**Figure 3 | Estimation of the proportion of lifetime cancer risk that is not due entirely to 'bad luck'. a, b,** Estimations based on total tissue stem-cell divisions originally reported in Tomasetti & Vogelstein<sup>5</sup> (**a**) and total tissue cell divisions (**b**). Red dots are cancers used to compute the 'intrinsic' risk

linear regression lines (red dashed lines). Blue dots are cancers known to have substantial extrinsic risks from epidemiology studies. The numbers in parentheses are the estimated percentages of cancer risks that are due to factors other than intrinsic risks.

small intestine, extranodal non-Hodgkin lymphoma, testicular, anal and anorectal cancers) have been steadily increasing, and their current incidences are substantially higher than their historical minima in the past 40 years<sup>28</sup> (Extended Data Fig. 1). Moreover, the mortality trend of lung cancer from 1930–2011 (ref. 29), which usually mirrors its incidence trend, shows a more than 15-fold increase for lung cancer risk. These substantial increases in incidence suggest that large risk proportions are attributable to changing environments (for example, smoking and air pollutants and their role in the risk of developing lung cancer). Collectively, nearly all major cancers have been covered in these epidemiological studies, further supporting the hypothesis of substantial extrinsic risks for most cancers. Notably, most of these cancers from the epidemiological and SEER results, except for small intestine, are located above the red 'intrinsic' risk lines in Fig. 3a, b (blue points). Accounting for the external factors would move them closer to the proposed 'intrinsic' line, further supporting the conjecture that the intrinsic line is mainly defined by cancers without compelling known epidemiological risk, whereas those above are at higher risks owing to extrinsic factors.

### Analysis of mutational signatures

In addition to epidemiological studies, we evaluated recent studies on mutational signatures in cancer. These are regarded as 'fingerprints' left on cancer genomes by different mutagenic processes<sup>30</sup>, revealing ~30 distinct signatures among various cancers<sup>31</sup>. Analysis of these signatures was therefore used to shed light on the proportion of intrinsic versus extrinsic origins of cancer. Two signature mutations, 1A/1B (see ref. 31), demonstrated strong positive correlations with age in the majority of cancers, suggesting that they are acquired at a relatively constant rate over the lifetime of cancer patients and thus probably result from intrinsic processes; however, all other signature mutations (~30) lack the consistent correlations with age, suggesting that they are acquired at different rates in life and thus are probably a consequence of extrinsic carcinogen exposures<sup>31</sup>. Indeed, several mutational signatures have been linked to known factors such as UV radiation and smoking<sup>31</sup>. We therefore categorized the signatures into intrinsic (type 1A/1B) and extrinsic mutations with known or unknown factors, and summarized their corresponding percentages in Extended Data Table 3. Notably, many cancers have substantial extrinsic mutations with



**Figure 4 | Theoretical lifetime intrinsic risks (tLIR) for cancers based on different number of hits ( $k$ ) required for cancer onset. a, b, The green (a) and blue (b) dashed lines are the ‘intrinsic’ risk lines estimated on the basis of total reported stem-cell numbers and total homeostatic tissue cells, respectively. The intrinsic stem-cell mutation rate ( $r$ ) is assumed to be**

known factors. More importantly, cancers known to have substantial environmental risk proportions, for example, breast cancer<sup>15</sup>, prostate cancer<sup>16</sup>, colorectal cancer<sup>18</sup>, melanoma<sup>19</sup>, head and neck cancer<sup>21</sup>, oesophageal cancer<sup>22</sup>, cervical cancer<sup>23</sup>, liver cancer<sup>26</sup> and stomach cancer<sup>27</sup>, all harbour large percentages of total extrinsic mutational signatures. This suggests that the percentages of total extrinsic mutational signatures can serve as a good surrogate for extrinsic cancer risks. While a few cancers have relatively large proportions of intrinsic mutations (>50%), the majority of cancers have large proportions of extrinsic mutations, for example, ~100% for myeloma, lung and thyroid cancers and ~80–90% for bladder, colorectal and uterine cancers, indicating substantial contributions of carcinogen exposures in the development of most cancers.

### Modelling theoretical lifetime intrinsic risk

Finally, in another independent model-driven approach to dissecting the risk contribution of the intrinsic processes, we modelled the potential lifetime cancer risk due to intrinsic stem-cell mutation errors by varying the number of hits (that is, driver gene mutations), denoted by  $k$ , required for cancer onset. We derived the probability distribution of the propagation of driver gene mutations from one generation to the next, and subsequently established the theoretical relationship between cell divisions and the degree of lifetime cancer risk due to intrinsic cell mutation errors alone, which we refer to as the theoretical lifetime intrinsic risk (tLIR). To overcome the limitation of inaccurate estimation in the reported stem-cell numbers<sup>5</sup>, we calculated tLIR using both the reported stem-cell number (tLIRsc) and the total tissue cell number (tLIRtt). The latter is equivalent to assuming all homeostatic tissue cells to be stem cells, representing an extreme overestimation of tissue stem cells, which consequently leads to a conservative estimation of the upper bounds in tLIR. The somatic mutation rate in tumours is estimated to be  $5 \times 10^{-10}$  per nucleotide site per cell division<sup>32–34</sup>. On this basis, in our initial calculation we used an intrinsic mutation rate ( $r$ ) of  $1 \times 10^{-8}$  per cell division, which is equivalent to approximately 20 mutable nucleotide sites for each driver gene where the driver gene will mutate if at least one site mutates. As shown in Fig. 4a, b, if only one hit (that is, mutation of one designated driver gene) is required to develop cancer—that is,  $k=1$ —the lifetime risk for almost all cancers is close to 100%. This confirms that one mutation is not enough for cancer onset (otherwise everyone would theoretically acquire each type of cancer). If two driver gene mutations are needed,  $k=2$ , the modelled intrinsic

$1 \times 10^{-8}$  per cell division. The red dashed lines are the ‘intrinsic’ risk lines estimated on the basis of the observed data using the same mechanism as Fig. 3a. Adjusted (adj.) basal and adjusted melanoma represent cancer risks after adjusting for the effect of sun exposure and UV radiation. AML, acute myeloid leukaemia.

risk becomes small for cancers with a small total number of stem-cell divisions; however, it is still very large for those with higher stem-cell divisions, and even unreasonably large for some cancers by surpassing the corresponding observed total lifetime cancer risks (adjusted basal cell carcinoma, colon adenocarcinoma, adjusted melanoma, small intestine cancer, acute myeloid leukaemia and duodenal cancer; Fig. 4a). It is therefore unlikely that, at least in these cancers, two hits will suffice to induce cancer. As shown in Fig. 4, if we consider the more reasonable case where three mutations are required<sup>35</sup>,  $k=3$ , almost all modelled intrinsic risks (both tLIRsc and tLIRtt) drop well below our earlier ‘intrinsic’ risk lines estimated conservatively from the observed data alone (red dashed lines, estimated based on observed data following the same mechanism as Fig. 3a). The lifetime risk drops even further for  $k=4$  and beyond. The extrinsic risks based on the tLIRsc and tLIRtt are further summarized in Extended Data Table 4. This modelling approach demonstrates that cancer risk due to intrinsic stem-cell mutation errors alone is low for almost all cancers that require over two mutations, indeed it is lower than the relatively conservative estimate based on data alone (red lines, Fig. 4). As the driver gene mutation rate in stem-cell division is a key parameter, we further conducted sensitivity analyses with different rates ( $r = 1 \times 10^{-10}$  to  $1 \times 10^{-6}$ ) to examine how this may affect the tLIR (Extended Data Figs 2 and 3). The results show that for  $k=3$ , when  $r < 1 \times 10^{-7}$  (~200 sites for each driver gene hit), almost all modelled intrinsic risks are below the observed ‘intrinsic’ risk line (red lines); when  $r = 1 \times 10^{-6}$  (~2,000 sites for each driver gene hit), the majority of modelled intrinsic risks are still well below the observed ‘intrinsic’ risk lines, particularly those with small total number of divisions (Extended Data Fig. 2). For  $k=4$ , when  $r < 1 \times 10^{-6}$ , almost all modelled intrinsic risks are below the observed ‘intrinsic’ risk lines estimated through the data-driven approach (Extended Data Fig. 3). These sensitivity analyses demonstrate that our conclusions are highly robust, and that the attribution of intrinsic mutations to lifetime cancer risk through stem-cell divisions, particularly for those cancers with low risk, is rather small, even using widely different intrinsic mutation rates.

In summary, we find that a simple regression analysis cannot distinguish between intrinsic and extrinsic factors. We have provided a new framework to quantify the lifetime cancer risks from both intrinsic and extrinsic factors on the basis of four independent approaches that are data-driven and model-driven, with and without using the stem-cell estimations. Importantly, these four approaches

provide a consistent estimate of contribution of extrinsic factors of >70–90% in most common cancer types. This is consistent with the overall conclusion regarding the role of extrinsic factors in cancer development.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 15 April; accepted 23 October 2015.**

**Published online 16 December 2015.**

- Sell, S. Stem cell origin of cancer and differentiation therapy. *Crit. Rev. Oncol. Hematol.* **51**, 1–28 (2004).
- Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105–111 (2001).
- Cairns, J. Mutation selection and the natural history of cancer. *Nature* **255**, 197–200 (1975).
- Visvader, J. E. Cells of origin in cancer. *Nature* **469**, 314–322 (2011).
- Tomasetti, C. & Vogelstein, B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78–81 (2015).
- Ashford, N. A. *et al.* Cancer risk: role of environment. *Science* **347**, 727 (2015).
- Wild, C. *et al.* Cancer risk: role of chance overstated. *Science* **347**, 728 (2015).
- Potter, J. D. & Prentice, R. L. Cancer risk: tumors excluded. *Science* **347**, 727 (2015).
- Gotay, C., Dummer, T. & Spinelli, J. Cancer risk: prevention is crucial. *Science* **347**, 728 (2015).
- Song, M. & Giovannucci, E. L. Cancer risk: many factors contribute. *Science* **347**, 728–729 (2015).
- O'Callaghan, M. Cancer risk: accuracy of literature. *Science* **347**, 729 (2015).
- Tomasetti, C. & Vogelstein, B. Cancer risk: accuracy of literature—response. *Science* **347**, 729–731 (2015).
- Altenberg, L. Statistical problems in a paper on variation in cancer risk among tissues, and new discoveries. Preprint at <http://arxiv.org/abs/1501.04605> (2015).
- Torre, L. A. *et al.* Global cancer statistics, 2012. *CA Cancer J. Clin.* **65**, 87–108 (2015).
- Gray, J., Evans, N., Taylor, B., Rizzo, J. & Walker, M. State of the evidence: the connection between breast cancer and the environment. *Int. J. Occup. Environ. Health* **15**, 43–78 (2009).
- Shimizu, H. *et al.* Cancers of the prostate and breast among Japanese and white immigrants in Los Angeles County. *Br. J. Cancer* **63**, 963–966 (1991).
- Haggard, F. A. & Boushey, R. P. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clin. Colon Rectal Surg.* **22**, 191–197 (2009).
- Johnson, I. T. & Lund, E. K. Review article: nutrition, obesity and colorectal cancer. *Aliment. Pharmacol. Ther.* **26**, 161–181 (2007).
- Parkin, D. M., Mesher, D. & Sasieni, P. 13. Cancers attributable to solar (ultraviolet) radiation exposure in the UK in 2010. *Br. J. Cancer* **105** (Suppl 2), S66–S69 (2011).
- Koh, H. K., Geller, A. C., Miller, D. R., Grossbart, T. A. & Lew, R. A. Prevention and early detection strategies for melanoma and skin cancer. Current status. *Arch. Dermatol.* **132**, 436–443 (1996).
- Blot, W. J. *et al.* Smoking and drinking in relation to oral and pharyngeal cancer. *Cancer Res.* **48**, 3282–3287 (1988).
- Kamangar, F., Chow, W.-H., Abnet, C. & Dawsey, S. Environmental causes of esophageal cancer. *Gastroenterology Clin. North Am.* **38**, 27–57 (2009).
- Bosch, F. X. *et al.* Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. *J. Natl. Cancer Inst.* **87**, 796–802 (1995).
- Frisch, M. *et al.* Sexually transmitted infection as a cause of anal cancer. *N. Engl. J. Med.* **337**, 1350–1358 (1997).
- Chaturvedi, A. K. *et al.* Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *J. Clin. Oncol.* **29**, 4294–4301 (2011).
- El-Serag, H. B. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology* **142**, 1264–1273 (2012).
- Helicobacter* and Cancer Collaborative Group. Gastric cancer and *Helicobacter pylori*: a combined analysis of 12 case control studies nested within prospective cohorts. *Gut* **49**, 347–353 (2001).
- Surveillance, Epidemiology, and End Results (SEER) Program. SEER\*Stat Database: Incidence—SEER 9 Regs Research Data, Nov 2014 Sub (1973–2012) (National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, 2015).
- Dela Cruz, C. S., Tanoue, L. T. & Matthay, R. A. Lung cancer: epidemiology, etiology, and prevention. *Clin. Chest Med.* **32**, 605–644 (2011).
- Alexandrov, L. B. & Stratton, M. R. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Curr. Opin. Genet. Dev.* **24**, 52–60 (2014).
- Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
- Jones, S. *et al.* Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl Acad. Sci. USA* **105**, 4283–4288 (2008).
- Tomasetti, C., Vogelstein, B. & Parmigiani, G. Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proc. Natl Acad. Sci. USA* **110**, 1999–2004 (2013).
- Bozic, I. & Nowak, M. A. Unwanted evolution. *Science* **342**, 938–939 (2013).
- Tomasetti, C., Marchionni, L., Nowak, M. A., Parmigiani, G. & Vogelstein, B. Only three driver gene mutations are required for the development of lung and colorectal cancers. *Proc. Natl Acad. Sci. USA* **112**, 118–123 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank L. Obeid for constructive comments. This work was supported in part by NCI grants 97132 and 168409 and Stony Brook NYSTEM award C026716.

**Author Contributions** Y.A.H. formulated the hypothesis. S.W. and Y.A.H. designed the research. S.W. and W.Z. performed mathematical and statistical analysis. S.W., S.P., W.Z. and Y.A.H. performed research. S.W., S.P., W.Z. and Y.A.H. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.W. (Song.Wu@stonybrook.edu) or Y.A.H. (Yusuf.Hannun@sbumed.org).



## METHODS

**Data reporting.** No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

**Derivation of the probability of possessing  $k$  hits after  $n$  cell divisions for one cell.** On the basis of the theory of the clonal stem-cell origin of cancer, in a given tissue the stem cell would first go through  $m$  rounds of symmetric divisions (for each division, each stem cell would divide into two daughter stem cells) to reach a total of  $S$  stem cells ( $S = 2^m$ ) at the steady state. Subsequently, these  $S$  stem cells would go through  $a$  rounds of asymmetric divisions (for each division, each stem cell would yield only one daughter stem cell) throughout the lifetime of the tissue. This means that the total rounds of lifetime stem-cell divisions per generation is  $n = m + a$ . Information on the total rounds of symmetric and asymmetric divisions as well as the total number of stem cells in the steady state for various tissues discussed in this work has been extracted from supplementary table 1 of Tomasetti & Vogelstein<sup>5</sup>. With  $k$  hits (mutations of  $k$  predetermined driver genes) on a stem cell required for cancer onset, the number of possible cell states of a given stem-cell generation would be  $k + 1$ , including a zero state with no hit. If we assume that once a hit occurs it cannot be reversed and therefore be carried to all progeny cells, then a cell state may only transition from lower to higher or equal levels from generation to generation. In Extended Data Fig. 4, we demonstrate with  $k = 3$  the state transitions of accumulating driver gene mutations. Let  $X_g$  denote the number of driver gene mutations accumulated at generation  $g$ , and  $r$  be the intrinsic driver gene mutation rate due to random errors during DNA replication; the transition probabilities to generation  $g + 1$  with  $i$  mutations from the previous generation  $g$  with  $j \leq i$  mutations are derived as follows:

$$P(X_{g+1} = i) = \sum_{j=0}^i P(X_{g+1} = i | X_g = j) P(X_g = j) \\ = \sum_{j=0}^i \binom{k-j}{i-j} r^{i-j} (1-r)^{k-i} P(X_g = j)$$

In particular, for the emission state  $i = 0$ :

$$P(X_{g+1} = 0) = (1-r)^k P(X_g = 0)$$

For the absorbing state  $i = k$ :

$$P(X_{g+1} = k) = \sum_{j=0}^k r^{k-j} P(X_g = j)$$

Based on these, the computing algorithm is derived as follows:  
Set the initial cell state at generation 0:

$$P(X_0 = 0) = 1; P(X_0 = 1) = 0; \dots; P(X_0 = k) = 0$$

For  $g = 1, \dots, n$  and  $0 \leq i \leq k$ , we compute the following probabilities iteratively:

$$P(X_g = i) = \sum_{j=0}^i \binom{k-j}{i-j} r^{i-j} (1-r)^{k-i} P(X_{g-1} = j)$$

where  $n$  is the total number of divisions that one stem cell may experience during its lifetime.

**Derivation of the theoretical lifetime intrinsic risk (tLIR) of cancer for a given tissue.** As mentioned previously, we assume stem cells in a specific tissue undergo two phases of divisions (Extended Data Fig. 5): (1) a total of  $m$  symmetric divisions before full tissue development, and (2) a total of  $a$  asymmetric divisions for normal tissue turnovers. So in a fully developed tissue, there is a total of  $S = 2^m$  stem cells. For each stem cell, the probability of possessing all  $k$  hits for cancer onset after  $n = m + a$  rounds of divisions is  $P(X_n = k)$ , which can be calculated from the previous part. Therefore, the theoretical lifetime intrinsic risk (tLIR) of developing cancer—that is, the probability of at least one stem cell containing  $k$  hits during its lifetime—can be expressed as:

$$\text{tLIR} = 1 - [1 - P(X_n = k)]^S$$

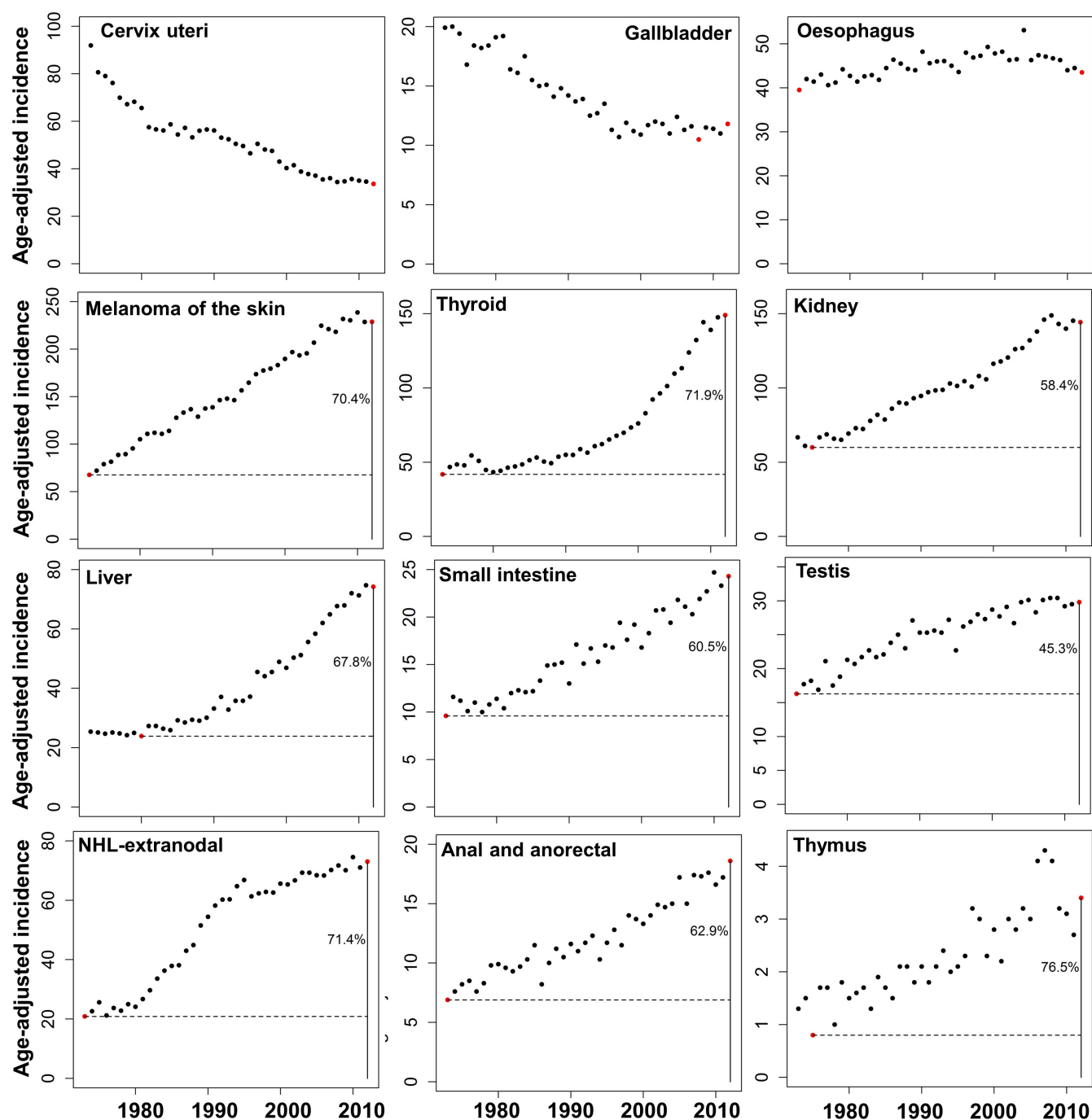
**Estimating cancer risk for different tissues.** The rounds of symmetric and asymmetric divisions for different tissues were adopted from supplementary table 1 of Tomasetti & Vogelstein<sup>5</sup>. In particular, the rounds of symmetric divisions,  $m$ , is equal to the integer part of  $\log_2 S$ , where  $S$  is the number of normal stem cells in the tissue of origin (data from ref. 5), and the rounds of asymmetric divisions  $a$  was the column labelled 'd' in supplementary table 1 of ref. 5. Sensitivity analyses have been conducted for scenarios with a broad range of mutation rates, from  $1 \times 10^{-10}$  to  $1 \times 10^{-6}$ , and several required hits ( $k = 1, 2, 3, 4$ ).

**Lower-bound estimates of extrinsic risks with the SEER data.** As a program of the National Cancer Institute (NCI), SEER (Surveillance, Epidemiology, and End Results Program) is a source of information on cancer incidence and survival in the USA (<http://seer.cancer.gov/>). The age-adjusted cancer incidences were extracted from the database 'SEER 9 Regs Research Data, Nov 2014 Sub (1973–2012) <Katrina/Rita Population Adjustment>' using the SEER\*Stat 8.2.1 (ref. 28). For several cancers, it has been observed that their incidence rates have increased markedly during the past 40 years (Extended Data Fig. 1). For these cancers, it is reasonable to assume that anything above the historical minimum incidence should be attributed to some environmental/extrinsic factors. Therefore, we can establish the following inequality:

Extrinsic risk  $> (1 - \text{historical minimum incidence rate/incidence rate in 2012})$ .

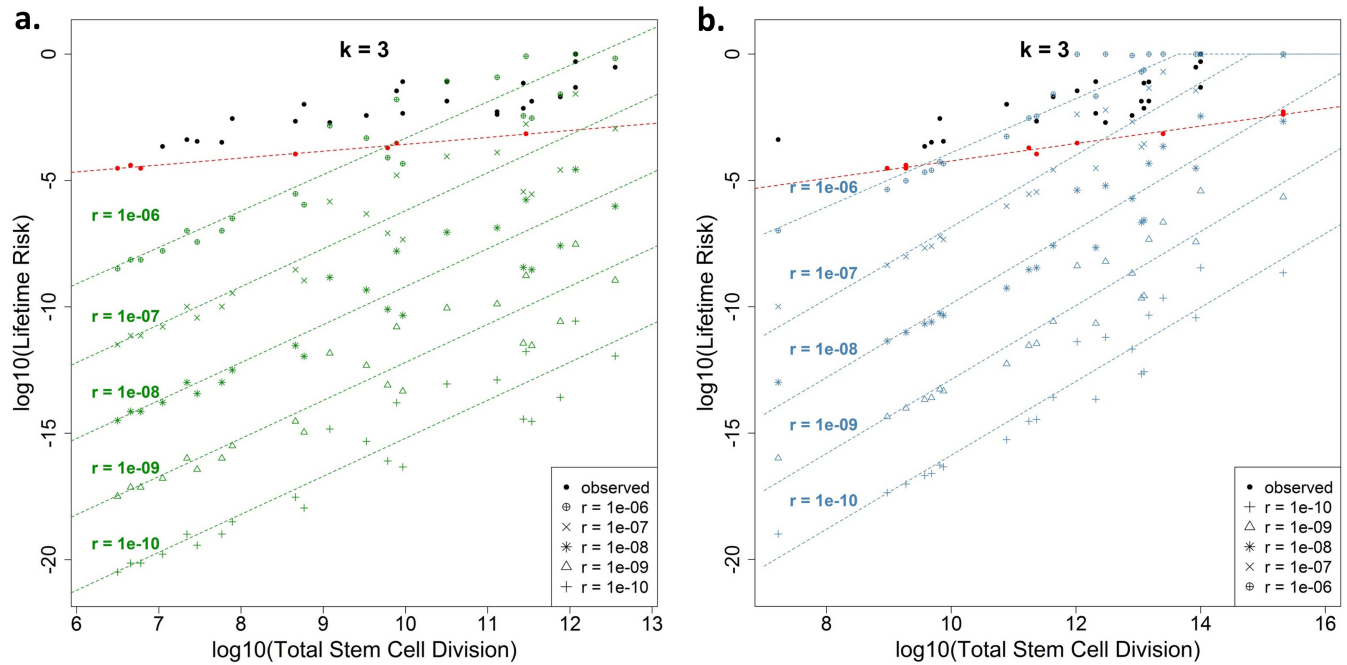
Correspondingly, the lower bounds of contributions by extrinsic factors for these cancers can be calculated. As shown in Extended Data Fig. 1, some cancers show substantial contributions from extrinsic factors.

**Data and statistical analysis.** The observed lifetime cancer risks and the cumulative number of divisions ( $n$ ) of all stem cells per lifetime are adopted from supplementary table 1 of Tomasetti & Vogelstein<sup>5</sup>. The total tissue cell divisions are from our evaluation of the data (Supplementary Information). For the robustness analysis of Fig. 3 as shown in Extended Data Table 1, error terms following the normal distribution with mean 0 and standard deviations of 1 or 0.4 were added to the  $\log_{10}(\text{total stem-cell division})$  or  $\log_{10}(\text{total cell division})$ . These allow the number of total stem-cell and cell divisions to vary approximately within a range of  $\sim 1/100$ –100-fold or  $\sim 1/5$ –5-fold, respectively. On the basis of the new data set with measurement errors, the excess risks for each cancer were quantified. This process is repeated 1,000 times, and from this the mean, the 2.5 and the 97.5 percentiles (namely the 95% confidence intervals) of the excess risk for each cancer are tabulated. In calculating the percentage of intrinsic versus extrinsic mutations based on mutational signatures from cancer genome, we define the intrinsic mutation as those with signatures 1A/1B, and extrinsic mutation as all other mutational signatures (2–21, R1–R3, U1 and U2). The corresponding data were obtained from supplementary figures 59–88 of ref. 31. All statistical analyses and mathematical calculations were performed using R (version 3.1.2).



**Extended Data Figure 1 | Examples of increased cancer incidence trends from 1973–2012 in SEER data.** The cancer types include melanoma, thyroid cancer, kidney cancer, liver cancer, small intestine cancer, testicular cancer, non-Hodgkin lymphoma (NHL), anal and anorectal cancer and thymus cancer. The horizontal dashed lines indicate the

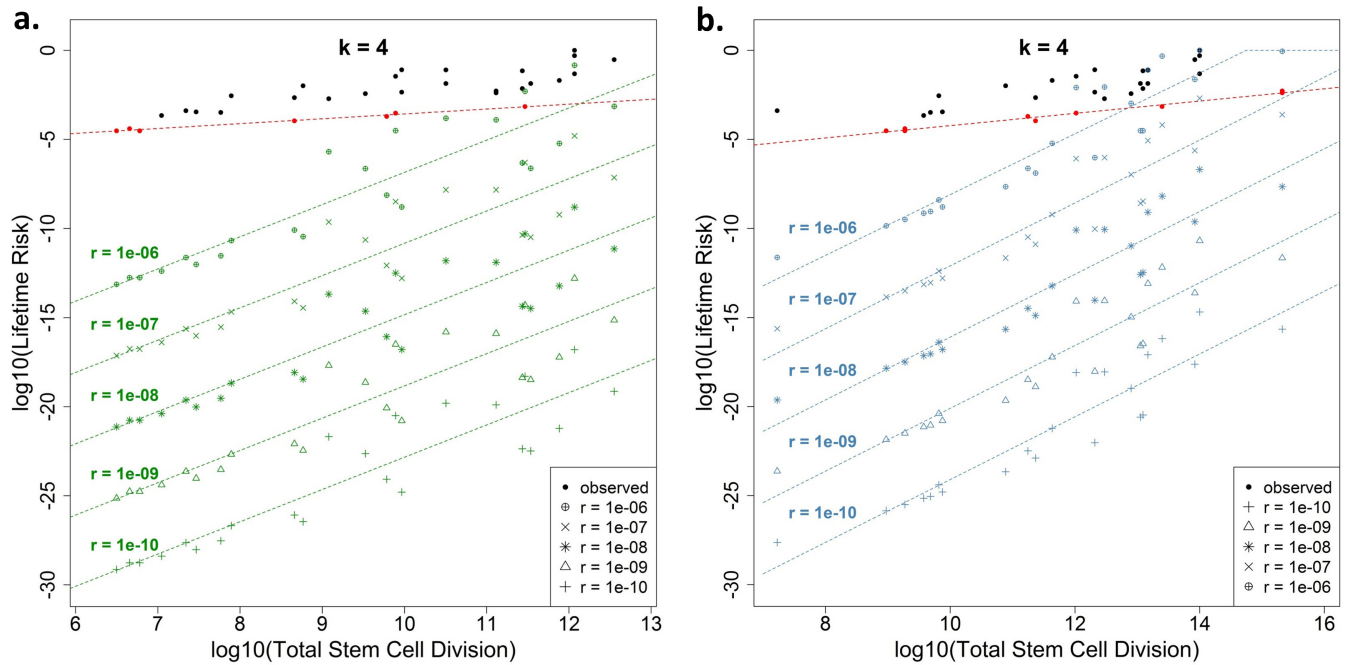
historical minimal incidence. The vertical solid lines indicate the most recent year. The numbers represent the minimal percentage of extrinsic risk. The cervix uteri cancer, gallbladder cancer and oesophageal cancer are examples with declining or consistent incidence trend. The incidence rate is per 100,000 people.



**Extended Data Figure 2 | Sensitivity analysis of different mutation rates on tLIR when the number of hits ( $k$ ) required is 3. a, b,** Theoretical intrinsic lifetime risks (tLIR) for cancers have been calculated based on five different mutation rates:  $r = 1 \times 10^{-10}$ ,  $1 \times 10^{-9}$ ,  $1 \times 10^{-8}$ ,  $1 \times 10^{-7}$ ,  $1 \times 10^{-6}$ . The red dashed lines are the ‘intrinsic’ risk lines based on the

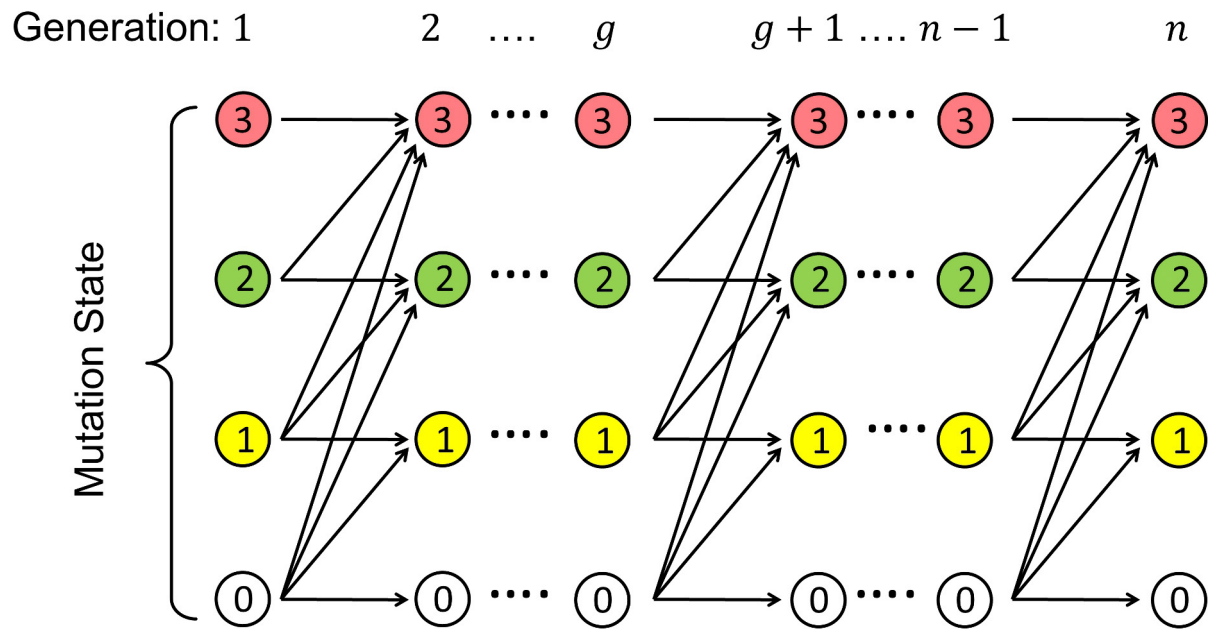
observed data following the same estimation mechanism as the intrinsic risk line in Fig. 3a. The green (a) and blue (b) dashed lines are the ‘intrinsic’ risk lines estimated based on total reported stem-cell numbers and total homeostatic tissue cells, respectively.



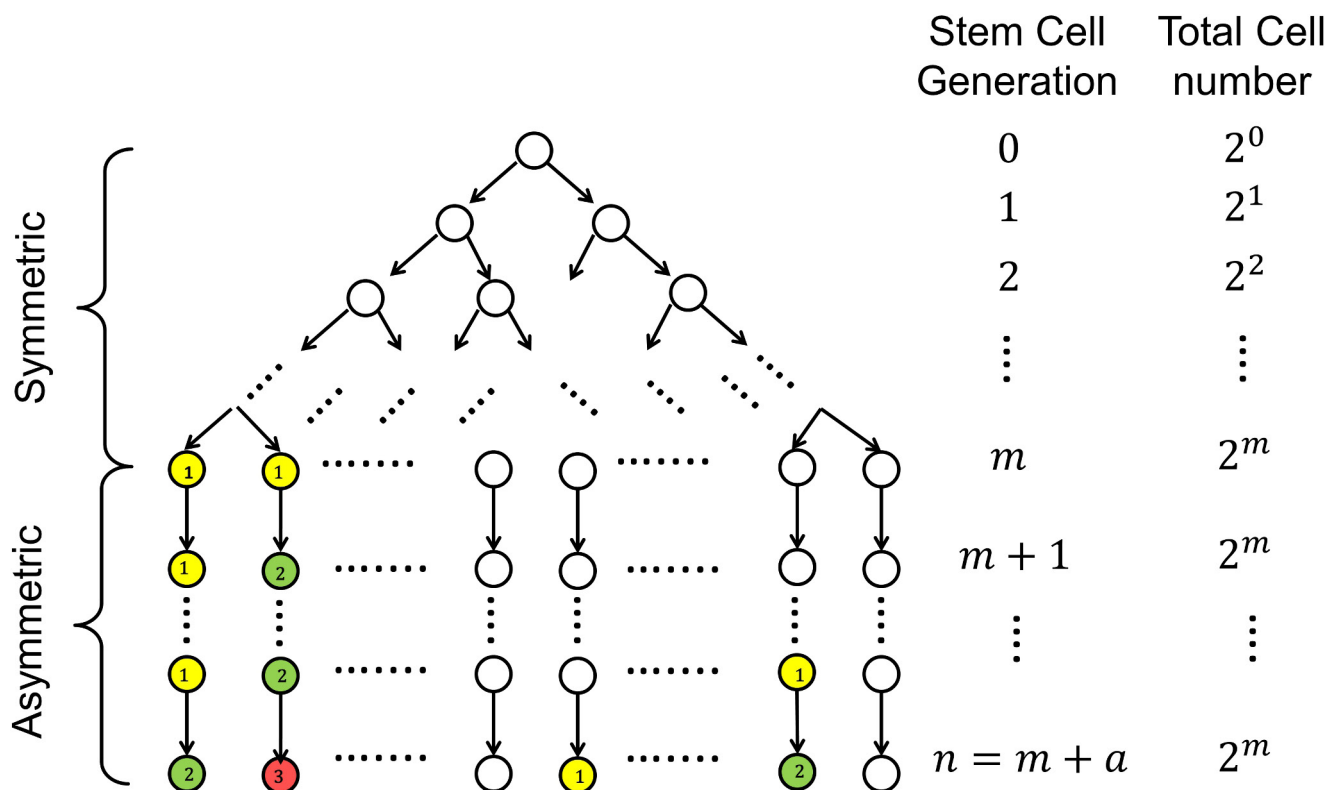


**Extended Data Figure 3 | Sensitivity analysis of different mutation rates on tLIR when the number of hits ( $k$ ) required is 4.** **a, b,** Theoretical intrinsic lifetime risks (tLIR) for cancers have been calculated based on five different mutation rates:  $r = 1 \times 10^{-10}$ ,  $1 \times 10^{-9}$ ,  $1 \times 10^{-8}$ ,  $1 \times 10^{-7}$ ,  $1 \times 10^{-6}$ . The red dashed lines are the 'intrinsic' risk lines based on the

observed data following the same estimation mechanism as the intrinsic risk line in Fig. 3a. The green (a) and blue (b) dashed lines are the 'intrinsic' risk lines estimated based on total reported stem-cell numbers and total homeostatic tissue cells, respectively.



**Extended Data Figure 4 | Intrinsic cancer risk modelling.** Part 1 of 2: propagation diagram of driver gene mutation states between generations in one stem cell, from which the stem-cell mutation transition probabilities from one generation to the next are computed.



**Extended Data Figure 5 | Intrinsic cancer risk modelling.** Part 2 of 2: schema of stem-cell divisions and driver gene mutations, from which the theoretical lifetime intrinsic risks (tLIR) for cancer due to  $k$  driver gene mutations are computed. Each coloured circle represents the mutation of a new driver gene in the given stem cell (yellow, first mutation; green,

second mutation; red, third mutation). If the mutation of 3 designated driver genes would induce a cancerous stem cell ( $k = 3$ ), then this diagram shows a cancer occurrence as the second stem cell in the last generation (generation  $n$ ) that has accumulated all 3 driver gene mutations.



Extended Data Table 1 | Robustness analysis on total stem-cell divisions and cell divisions estimates in Fig. 3

Name	Observed Risk	Total stem-cell divisions (Fig. 3A)			Total cell divisions (Fig. 3B)		
		Log10 (divisions)	Excess risk	Excess risk 95% CI*	Log10 (divisions)	Excess risk	Excess risk 95% CI*
AML	0.0041	11.11	>0.871	[0.623 , 0.962]	NA	NA	NA
Basal cell	0.3	12.55	>0.996	[0.985 , 0.999]	14.42	>0.995	[0.99 , 0.998]
Breast	0.123	NA	NA	NA	14.54	>0.987	[0.974 , 0.994]
CLL	0.0052	11.11	>0.899	[0.701 , 0.973]	NA	NA	NA
COAD	0.048	12.07	>0.980	[0.934 , 0.995]	14.40	>0.971	[0.943 , 0.986]
FAP COAD	1	12.07	>0.999	[0.997 , 1.000]	14.40	>0.999	[0.997 , 0.999]
Lynch COAD	0.5	12.07	>0.998	[0.994 , 1.000]	14.40	>0.997	[0.994 , 0.999]
Duodenum <sup>†</sup>	3.00E-04	9.89	-	-	NA	NA	NA
FAP Duodenum	0.035	9.89	>0.993	[0.980 , 0.998]	NA	NA	NA
Esophageal	0.00194	9.08	>0.906	[0.748 , 0.975]	NA	NA	NA
Gallbladder	0.0028	7.89	>0.967	[0.922 , 0.991]	NA	NA	NA
Glioblastoma	0.00219	8.43	>0.943	[0.868 , 0.984]	NA	NA	NA
Head & neck	0.0138	10.50	>0.973	[0.921 , 0.992]	NA	NA	NA
HPV Head & neck	0.07935	10.50	>0.995	[0.985 , 0.999]	NA	NA	NA
Hepatocellular	0.0071	11.43	>0.906	[0.720 , 0.975]	13.41	>0.932	[0.872 , 0.969]
HCV Hepatocellular	0.071	11.43	>0.991	[0.969 , 0.998]	13.41	>0.993	[0.986 , 0.997]
Lung (nonsmoker) <sup>‡</sup>	0.0045	9.97	>0.938	[0.835 , 0.982]	15.2	-	-
Lung (smoker)	0.081	9.97	>0.997	[0.990 , 0.999]	15.20	>0.958	[0.905 , 0.982]
Medulloblastoma <sup>†</sup>	0.00011	8.43	-	-	NA	NA	NA
Melanoma	0.0203	11.88	>0.960	[0.872 , 0.990]	NA	NA	NA
Osteosarcoma	0.00035	7.47	>0.790	[0.459 , 0.947]	11.79	>0.762	[0.568 , 0.887]
Arms osteosarcoma <sup>†,‡</sup>	4.00E-05	6.66	-	-	10.99	-	-
Head osteosarcoma <sup>†,‡</sup>	3.02E-05	6.78	-	-	11.1	-	-
Legs osteosarcoma	0.00022	7.05	>0.727	[0.306 , 0.930]	11.37	>0.761	[0.537 , 0.889]
Pelvis osteosarcoma <sup>†,‡</sup>	3.00E-05	6.50	NA	NA	10.81	-	-
Ovarian germ cell	0.000411	7.34	>0.832	[0.573 , 0.958]	NA	NA	NA
Pancreatic ductal	0.013589	11.54	>0.948	[0.805 , 0.987]	NA	NA	NA
Pancreatic islet <sup>†</sup>	0.000194	9.78	-	-	NA	NA	NA
Prostate	0.14	NA	NA	NA	11.81	>0.999	[0.999 , 1]
Small intestine <sup>†,‡</sup>	7.00E-04	11.47	-	-	14.22	-	-
Testicular	0.0037	9.53	>0.942	[0.843 , 0.984]	13.02	>0.914	[0.835 , 0.959]
Thyroid follicular	0.01026	8.77	>0.986	[0.964 , 0.996]	NA	NA	NA
Thyroid medullary	0.000324	7.77	>0.731	[0.308 , 0.928]	NA	NA	NA

Measurement errors were added to log<sub>10</sub>(divisions) and 1,000 simulations were carried out to calculate the mean and 95% confidence interval (CI) of the excess risks. See Methods for details. NA: data not available.

\*Confidence interval.

†Cancers used to compute the 'intrinsic' risk line based on total stem-cell divisions.

‡Cancers used to compute the 'intrinsic' risk line based on total cell divisions.

Extended Data Table 2 | Epidemiological studies on the extrinsic risks of various cancers

Cancer Types	Extrinsic risk	Examples of potential extrinsic risk factors*
Breast	substantial	Oral contraceptive, hormone replacement therapy, lifestyle (diet, smoking, alcohol, weight)
Prostate	substantial	Diet, obesity, smoking
Lung	>90%	Smoking; air pollutant
Colorectal	>75%	Diet, smoking, alcohol, obesity
Melanoma	65-86%	Sun exposure
Basal cell	~90%	UV
Hepatocellular	~80%	HBV, HCV
Gastirc	65-80%	H. pylori
Cervical	~90%	HPV
Head & Neck	~75%	Tobacco, alcohol
Esophageal	>75%	Smoking, alcohol, obesity, diet
Oropharyngeal	~70%	HPV
Thyroid	>72%	Diet low in iodine, radiation
Kidney	>58%	Smoking, obesity, workplace exposures
Thymus	>77%	Largely unclear
Small intestine	>61%	Diet, smoking, alcohol
Extranodal non-Hodgkin's lymphoma (NHL)	>71%	Chemicals, radiation, immune system deficiency
Testis	>45%	Largely unclear
Anal and anorectal cancers	>63%	HPV, smoking

\*<http://www.cancer.org/cancer>.

Extended Data Table 3 | Percentages of intrinsic versus extrinsic MS with known and unknown causes in different cancer types

	Intrinsic MS	Extrinsic MS - Known	Extrinsic MS - Unknown	Extrinsic MS - Total
ALL	65.8	34.2	0	34.2
AML	100	0	0	0
Bladder	14.2	71.2	14.6	85.8
Breast	35.5	60.1	4.4	64.5
Cervical	25.3	74.7	0	74.7
CLL	76.7	23.3	0	23.3
Colorectal	17.1	66	16.9	82.9
Esophageal	48	25.3	26.7	52
Glioblastoma	53.8	0	46.2	46.2
Glioma-Low Grade	9.2	2.8	88	90.8
Head & Neck	24.9	75.1	0	75.1
Kidney Chromophobe	17.4	37.5	45.1	82.6
Kidney Clear Cell	66.5	4.1	29.4	33.5
Kidney Papillary	0	15.7	84.3	100
Liver	10.9	21.3	67.8	89.1
Lung Adenocarcinoma	9.1	73.8	17.1	90.9
Lung - Small Cell	0	92.8	7.2	100
Lung-Squamous	0	47	53	100
Lymphoma B-cell	46.3	33.4	20.3	53.7
Medulloblastoma	48.4	0	51.6	51.6
Melanoma	7.2	90.9	1.9	92.8
Myeloma	0	19.9	80.1	100
Neuroblastoma	53.2	0	46.8	46.8
Ovarian	36.6	63.4	0	63.4
Pancreatic	49.9	50.1	0	50.1
Pilocytic Astrocytoma	82.5	0	17.5	17.5
Prostate	32.2	10.2	57.6	67.8
Stomach	22.3	6.1	71.6	77.7
Thyroid	0	39.7	60.3	100
Uterine	10.7	65.5	23.8	89.3

Intrinsic mutational signatures (MS) includes signatures 1A/B, and extrinsic MS includes signatures 2–21, R1–R3, U1 and U2, excluding signature 11 for Temozolomide, an alkylating agent used for chemotherapy. The blue, yellow and red colours highlight cancers that are have substantial extrinsic risk proportions based on epidemiological data, MS with known causes and MS with unknown causes, respectively. Data from the supplementary figs 59–88 in ref. 31.



**Extended Data Table 4 | Percentages of extrinsic risks based on the reported stem-cell estimates and total homeostatic tissue cells, as shown in Fig. 4**

Extrinsic Risks	Based on stem cell estimates				Based on total homeostatic tissue cells			
Cancer Type	k=1	k=2	k=3	k=4	k=1	k=2	k=3	k=4
AML	H.T.O.	H.T.O.	1.000	1.000	H.T.O.	H.T.O.	0.465	1.000
Basal cell	H.T.O.	0.462	1.000	1.000	H.T.O.	H.T.O.	1.000	1.000
CLL	H.T.O.	H.T.O.	1.000	1.000	H.T.O.	H.T.O.	0.578	1.000
COAD	H.T.O.	H.T.O.	0.999	1.000	H.T.O.	H.T.O.	0.928	1.000
FAP COAD	H.T.O.	0.630	1.000	1.000	H.T.O.	H.T.O.	0.997	1.000
Lynch COAD	H.T.O.	0.260	1.000	1.000	H.T.O.	H.T.O.	0.993	1.000
Duodenum	H.T.O.	H.T.O.	1.000	1.000	H.T.O.	H.T.O.	0.986	1.000
FAP Duodenum	H.T.O.	0.977	1.000	1.000	H.T.O.	H.T.O.	1.000	1.000
Esophageal	H.T.O.	0.946	1.000	1.000	H.T.O.	H.T.O.	0.997	1.000
Gallbladder	H.T.O.	1.000	1.000	1.000	H.T.O.	0.974	1.000	1.000
Glioblastoma	H.T.O.	0.995	1.000	1.000	H.T.O.	H.T.O.	1.000	1.000
Head & neck	H.T.O.	0.631	1.000	1.000	H.T.O.	H.T.O.	0.997	1.000
HPV Head & neck	H.T.O.	0.936	1.000	1.000	H.T.O.	H.T.O.	0.999	1.000
Hepatocellular	H.T.O.	0.572	1.000	1.000	H.T.O.	H.T.O.	1.000	1.000
HCV Hepatocellular	H.T.O.	0.957	1.000	1.000	H.T.O.	H.T.O.	1.000	1.000
Lung (nonsmoker)	H.T.O.	0.971	1.000	1.000	H.T.O.	H.T.O.	1.000	1.000
Lung (smoker)	H.T.O.	0.998	1.000	1.000	H.T.O.	0.388	1.000	1.000
Medulloblastoma	H.T.O.	0.904	1.000	1.000	H.T.O.	H.T.O.	1.000	1.000
Melanoma	H.T.O.	0.444	1.000	1.000	H.T.O.	0.444	1.000	1.000
Osteosarcoma	H.T.O.	1.000	1.000	1.000	H.T.O.	0.624	1.000	1.000
Arms osteosarcoma	H.T.O.	0.999	1.000	1.000	H.T.O.	0.269	1.000	1.000
Head osteosarcoma	H.T.O.	0.999	1.000	1.000	H.T.O.	0.032	1.000	1.000
Legs osteosarcoma	H.T.O.	1.000	1.000	1.000	H.T.O.	0.718	1.000	1.000
Pelvis osteosarcoma	H.T.O.	1.000	1.000	1.000	H.T.O.	0.542	1.000	1.000
Ovarian germ cell	H.T.O.	0.999	1.000	1.000	H.T.O.	0.999	1.000	1.000
Pancreatic ductal	H.T.O.	0.806	1.000	1.000	H.T.O.	H.T.O.	1.000	1.000
Pancreatic islet	H.T.O.	0.611	1.000	1.000	H.T.O.	H.T.O.	1.000	1.000
Small intestine	H.T.O.	H.T.O.	0.998	1.000	H.T.O.	H.T.O.	0.684	1.000
Testicular	H.T.O.	0.973	1.000	1.000	H.T.O.	H.T.O.	0.999	1.000
Thyroid follicular	H.T.O.	1.000	1.000	1.000	H.T.O.	0.866	1.000	1.000
Thyroid medullary	H.T.O.	0.999	1.000	1.000	H.T.O.	0.785	1.000	1.000

Extrinsic risk =  $1 - (tLIRsc \text{ or } tLIRtt) / \text{observed risk}$ . H.T.O., higher than the observed.

# SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination

Dorothy Yanling Zhao<sup>1,2,3</sup>, Gerald Gish<sup>2\*</sup>, Ulrich Braunschweig<sup>1\*</sup>, Yue Li<sup>1,4</sup>, Zuyao Ni<sup>1</sup>, Frank W. Schmitges<sup>1</sup>, Guoqing Zhong<sup>1</sup>, Ke Liu<sup>5</sup>, Weiguo Li<sup>5</sup>, Jason Moffat<sup>1,3</sup>, Masoud Vedadi<sup>5</sup>, Jinrong Min<sup>5</sup>, Tony J. Pawson<sup>2,3</sup>, Benjamin J. Blencowe<sup>1,3</sup> & Jack F. Greenblatt<sup>1,3</sup>

The carboxy-terminal domain (CTD) of the RNA polymerase II (RNAP II) subunit POLR2A is a platform for modifications specifying the recruitment of factors that regulate transcription, mRNA processing, and chromatin remodelling. Here we show that a CTD arginine residue (R1810 in human) that is conserved across vertebrates is symmetrically dimethylated (me2s). This R1810me2s modification requires protein arginine methyltransferase 5 (PRMT5) and recruits the Tudor domain of the survival of motor neuron (SMN, also known as GEMIN1) protein, which is mutated in spinal muscular atrophy. SMN interacts with senataxin, which is sometimes mutated in ataxia oculomotor apraxia type 2 and amyotrophic lateral sclerosis. Because POLR2A R1810me2s and SMN, like senataxin, are required for resolving RNA–DNA hybrids created by RNA polymerase II that form R-loops in transcription termination regions, we propose that R1810me2s, SMN, and senataxin are components of an R-loop resolution pathway. Defects in this pathway can influence transcription termination and may contribute to neurodegenerative disorders.

The CTD of POLR2A contains 52 heptapeptide repeats. The amino-terminal half of the CTD comprises repeats that mostly conform to the consensus Tyr1–Ser2–Pro3–Thr4–Ser5–Pro6–Ser7, whereas the C-terminal half consists of repeats that generally deviate from this consensus<sup>1</sup>. These repeats can be phosphorylated on Tyr1, Thr4, and all three serine residues, and specific CTD phosphorylation patterns are important for various aspects of chromatin regulation, transcription, and co-transcriptional RNA processing<sup>2–6</sup>. Two non-consensus human CTD arginine residues, R1603 and R1810, are conserved in vertebrates. It was found recently that asymmetric dimethylation (me2a) of R1810 by the CARM1 (also known as PRMT4) methyltransferase inhibits the expression of small nuclear RNA (snRNA) and small nucleolar RNA (snoRNA) genes in human cells<sup>7</sup>. It was also shown that this R1810 me2a mark can be bound *in vitro* by the Tudor domain of TDRD3 (ref. 7). At the c-Myc promoter, asymmetric histone arginine dimethylation by PRMT1 and CARM1 recruits TDRD3 and TOP3B to suppress R-loop accumulation<sup>8</sup>. We now show that R1810 can be symmetrically dimethylated, a modification requiring PRMT5. This R1810me2s modification recruits SMN, which then interacts with senataxin, a helicase needed for resolving R-loops in transcriptional termination regions.

## R1810me2s on the RNAP II CTD

Motivated initially by the question of whether the R1810me2a modification is recognized by TDRD3 *in vivo*, we performed immunoprecipitation analysis using tagged TDRD3 and the RNAP II POLR2D subunit. We observed that both tagged proteins could co-immunoprecipitate me2a-modified POLR2A, as detected by western blotting with the ASYMM24 antibody specific for Arg-me2a. In contrast, only POLR2D, and not TDRD3, co-immunoprecipitated a form

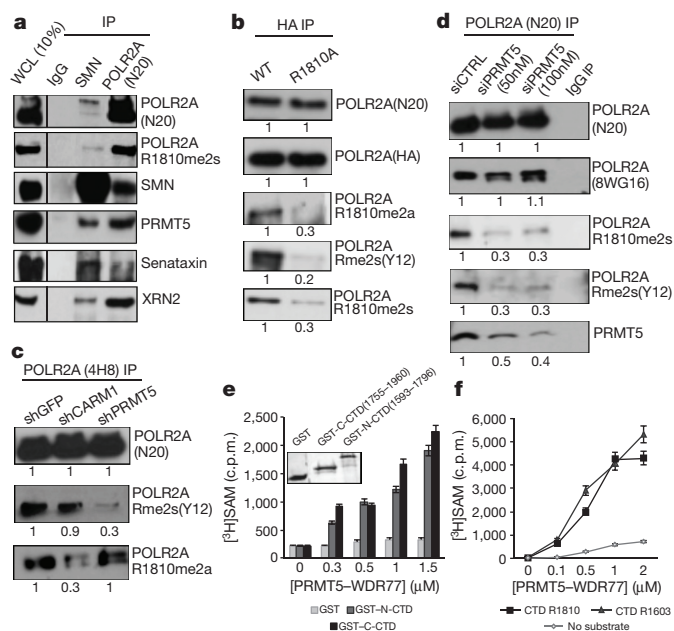
of POLR2A with an me2s modification that could be detected by SYMM10 or Y12 antibodies specific for Arg-me2s (Extended Data Fig. 1a). To determine whether R1810 is indeed dimethylated symmetrically, as well as asymmetrically, we generated polyclonal antibodies against an R1810me2s-containing 7-mer peptide and found that immunoprecipitated POLR2A is recognized by this R1810me2s antibody (Fig. 1a). To determine whether the Arg-me2s modification involved R1810, Raji cells stably expressing  $\alpha$ -amanitin-resistant, HA-tagged, wild-type or R1810A mutant POLR2A were generated<sup>7</sup>. After treatment with  $\alpha$ -amanitin to deplete endogenous  $\alpha$ -amanitin-sensitive RNAP II, followed by immunoprecipitation of RNAP II with anti-HA antibody, western blotting with antibodies recognizing R1810me2a<sup>7</sup> or R1810me2s, as well as with Y12 antibody, revealed that the R1810A mutation results in the loss of both R1810me2a and R1810me2s modifications (Fig. 1b and Extended Data Fig. 1b). The precipitated RNAP II was dephosphorylated before western blotting to enable more sensitive detection of R1810me2s (Extended Data Fig. 1c). The Y12 and R1810me2s antibodies recognize an Arg-me2s peptide bracketing CTD R1810 much better than peptides with no arginine modification or R1810me2a (see slot blots of Extended Data Fig. 1d). Therefore, R1810 is symmetrically dimethylated in cell extracts. Extended Data Fig. 1a also shows that TDRD3 recognizes R1810me2a in cell extracts, as well as *in vitro*<sup>7</sup>, although TDRD3 does not mediate inhibition of snRNA and snoRNA gene expression by R1810me2a<sup>9</sup>.

## R1810me2s modification requires PRMT5

Although PRMT9 can symmetrically dimethylate certain substrates, PRMT5 is the predominant methyltransferase that symmetrically dimethylates arginine in human cells<sup>10,11</sup>. As the PRMT5–WDR77

<sup>1</sup>Donnelly Centre, University of Toronto, Toronto, Ontario M5S 3E1, Canada. <sup>2</sup>Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada. <sup>3</sup>Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. <sup>4</sup>Department of Computer Science, University of Toronto, Toronto, Ontario M5S 3G4, Canada. <sup>5</sup>Structural Genomics Consortium, University of Toronto, Toronto, Ontario M5G 1L7, Canada.

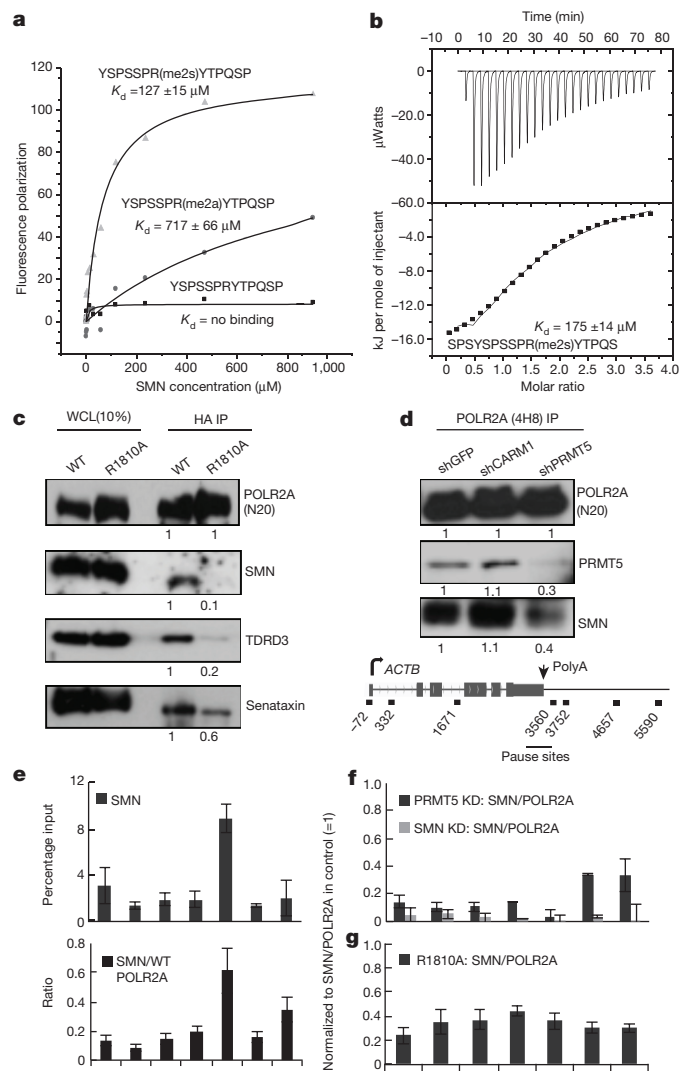
\*These authors contributed equally to this work.



**Figure 1 | Symmetric dimethylation of R1810 on the RNAP II CTD and a requirement for PRMT5.** **a–d**, Western blots, with relative quantifications underneath (in **b–d**), of immunoprecipitations (IP) from HEK293 (**a**, **c**, **d**) or Raji (**b**) whole-cell lysates (WCL) after stable (sh) or transient (small interfering RNA (si)) knockdowns (**c**, **d**). Anti-HA was used to precipitate HA-tagged POLR2A in **b**. **e**, **f**, Quantification of methylation of GST–CTD fusion proteins shown in the Coomassie-stained SDS gel of the insert (**e**), or 13-mer CTD peptides (**f**) with various concentrations of recombinant PRMT5–WDR77. Error bars denote s.e.m. ( $n = 3$  biological replicates).

complex associates with RNAP II through the CTD phosphatase FCP1 (ref. 12), which is consistent with our observation that PRMT5 co-purifies with RNAP II (Fig. 1a), we tested whether PRMT5 might be needed to symmetrically dimethylate R1810. HEK293 cell lines stably expressing shRNAs for CARM1, PRMT5, and GFP were generated, and endogenous RNAP II was precipitated. Western blotting revealed that CARM1 knockdown causes loss of the R1810me2a, but not the R1810me2s mark on POLR2A, whereas PRMT5 knockdown causes loss of the R1810me2s, but not the R1810me2a mark (Fig. 1c and Extended Data Fig. 1e, f). Transient siRNA-mediated knockdown of PRMT5 consistently also reduced R1810me2s, whereas overexpression of Flag-tagged PRMT5 increased R1810me2s (Fig. 1d and Extended Data Fig. 1g). These experiments indicated that PRMT5 is required *in vivo* for R1810me2s modification of the RNAP II CTD.

PRMT5 is the catalytic subunit of the methyltransferase, which also contains WDR77 (also known as MEP50)<sup>13,14</sup>. To test whether PRMT5 can directly methylate CTD arginine residues, we incubated recombinant PRMT5–WDR77 with tritiated S-adenosyl methionine ([<sup>3</sup>H]SAM) and recombinant GST–N-CTD, which contains CTD repeats 1–29 and includes R1603, or GST–C-CTD, which contains CTD repeats 24–52 and includes R1810 (GST was linked to the N-terminus of the CTD fragment in both cases). Scintillation counting was then used to monitor [<sup>3</sup>H]SAM labelling following glutathione–agarose pull-down of the GST fusion proteins. Both GST–C-CTD and GST–N-CTD were methylated above the background (GST alone, which contains 14 arginines) (Fig. 1e). When the PRMT5 methylation assays were repeated with biotinylated 13-mer peptides containing R1810 or R1603, methylation was again observed on both R1603 and R1810 (Fig. 1f). Therefore, PRMT5–WDR77 is needed to symmetrically dimethylate R1810 *in vivo* and can directly methylate R1603 and R1810 *in vitro*, but may require an additional co-factor<sup>12</sup> and/or appropriate CTD phosphorylation to specifically methylate R1810.



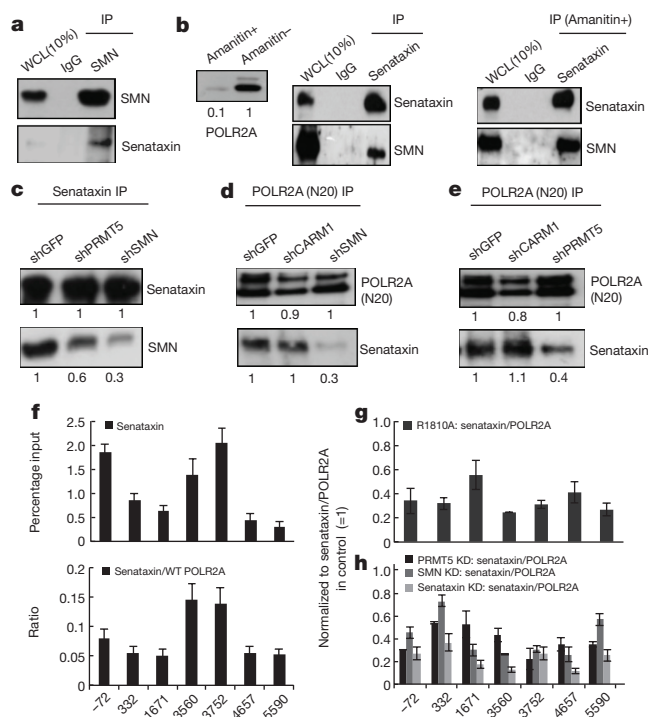
**Figure 2 | R1810me2s in the RNAP II CTD is recognized by SMN.**

**a**, Fluorescence polarization assays showing binding of SMN to FITC-labelled CTD peptides containing R1810. **b**, Isothermal titration calorimetry assays showing binding of SMN to a CTD peptide containing R1810me2s. **c**, **d**, Western blots, with relative quantifications underneath, of immunoprecipitations (IP) from HEK293 (**d**) or Raji (**c**) whole-cell lysates after various stable (sh) knockdowns (**d**). Anti-HA was used to precipitate HA-tagged POLR2A in **c**. **e–g**, Quantification of quantitative (qPCR) data from ChIP experiments at the indicated ACTB primer positions in HEK293 (**e**, **f**) or Raji (**g**) cells, along with effects of knocking down (KD) PRMT5 or SMN (**f**) or mutating R1810 to alanine (**g**). Error bars denote s.e.m. ( $n = 3$  biological replicates).

### Recognition of R1810me2s by SMN

Modified CTD residues and dimethylated arginine residues usually mediate their biological effects via interacting proteins, in the latter case, Tudor domain proteins<sup>15,16</sup>. Therefore, we sought a protein that recognizes R1810me2s. Because the Tudor domains of SMN and SPF30 (also known as SMNDC1), as well as those of TDRD1, TDRD2, TDRD9, and TDRD11, specifically bind Arg-me2s<sup>15</sup>, we used them in fluorescence polarization assays to identify whether any could bind an FITC-tagged 13-mer CTD peptide containing R1810me2s. Of these, only the Tudor domain of SMN exhibited binding, with much lower affinity for R1810me2a and R1603me2s peptides and no detectable affinity for the unmodified peptides (Fig. 2a and Extended Data Fig. 2a, c). In contrast, the TDRD3 Tudor domain showed weak affinity only for R1810me2a > R1603me2a above background (no modification or Arg-me2s), consistent with published data (data not shown)<sup>7,17</sup>.





**Figure 3 | R1810me2s and SMN recruit senataxin to RNAP II.**

**a–e**, Western blots, with relative quantification underneath, of immunoprecipitations (IP) from HEK293 whole-cell lysates after stable (sh) knockdowns of the indicated proteins (**c–e**). Cells were grown for 3 days in the presence of  $\alpha$ -amanitin in the right panel of **b** to eliminate endogenous POLR2A (western blot of left panel). **f–h**, Quantification of qPCR data from ChIP experiments at the indicated *ACTB* primer positions (see Fig. 2) in HEK293 (**f**, **h**) or Raji (**g**) cells, along with effects of knocking down (KD) PRMT5, SMN or senataxin (**h**) or mutating R1810 to alanine (**g**). Error bars denote s.e.m. ( $n = 3$  biological replicates).

Compared to a CTD peptide with R1810me2s alone, the presence of additional phospho-Tyr1 or phospho-Ser2 modifications, or both, on the peptide only slightly enhanced its binding to SMN in fluorescence polarization assays and had no significant effect in isothermal titration calorimetry assays (Fig. 2b and Extended Data Fig. 2b, d). Other phosphorylations near R1810 also had no effect on SMN binding *in vitro* (data not shown), indicating that the association of SMN with R1810me2s is not greatly influenced by CTD phosphorylation.

Using coimmunoprecipitation, we found that SMN and POLR2A interact (Fig. 1a). Consistent with specific recognition of the R1810me2s modification by SMN, immunoprecipitation of SMN from HEK293 cell extracts co-precipitated endogenous POLR2A with the R1810me2s modification (Fig. 1a). To test whether R1810 is important for the association of RNAP II with SMN *in vivo*, HA-tagged wild-type or mutant (R1810A) POLR2A was immunoprecipitated with anti-HA antibody. Western blotting showed that the R1810A mutation disrupts the association of RNAP II with both SMN and TDRD3 (Fig. 2c). As expected, immunoprecipitation of endogenous RNAP II from HEK293 cells expressing shRNAs for GFP, CARM1 or PRMT5 revealed that only knockdown of PRMT5 reduced co-precipitation of SMN (Fig. 2d). Similarly, transient siRNA-mediated knockdown of PRMT5 reduced the RNAP II–SMN interaction, whereas overexpression of PRMT5 enhanced it (Extended Data Fig. 3a, c).

The experiments described above reveal that R1810me2s is important for the SMN–RNAP II interaction. To examine whether SMN also displays R1810me2s-dependent association with RNAP II during transcription, SMN chromatin immunoprecipitation (ChIP) was performed. As shown in Fig. 2e and Extended Data Fig. 4a, SMN associates with the  $\beta$ -actin (*ACTB*) and *GAPDH* genes from their promoter regions to their termination regions. Moreover, PRMT5 knockdown

with shRNA (Fig. 2f and Extended Data Fig. 4b) or siRNA (not shown) strongly reduced the SMN ChIP signals along these genes, as did shRNA for SMN (Fig. 2f), showing the specificity of the SMN antibody for ChIP. The SMN ChIP signals on *ACTB* and *GAPDH* decreased upon expression of the POLR2A (R1810A) mutant, compared to wild type, after  $\alpha$ -amanitin treatment to eliminate endogenous POLR2A (Fig. 2g and Extended Data Fig. 4c). These experiments indicated that R1810me2s recruits SMN to RNAP II elongation complexes. Consistent with our observation that SMN is recruited from promoters to 3' ends, we found that immunoprecipitation of SMN co-precipitated RNAP II containing CTD phosphorylation on Ser2 and Ser5, modifications that are associated with RNAP II complexes engaged in 5'-end formation and elongation, respectively (Extended Data Fig. 3b). SMN ChIP-seq showed SMN occupancy only at promoter and termination regions, probably because SMN does not bind to DNA directly (Extended Data Fig. 3d).

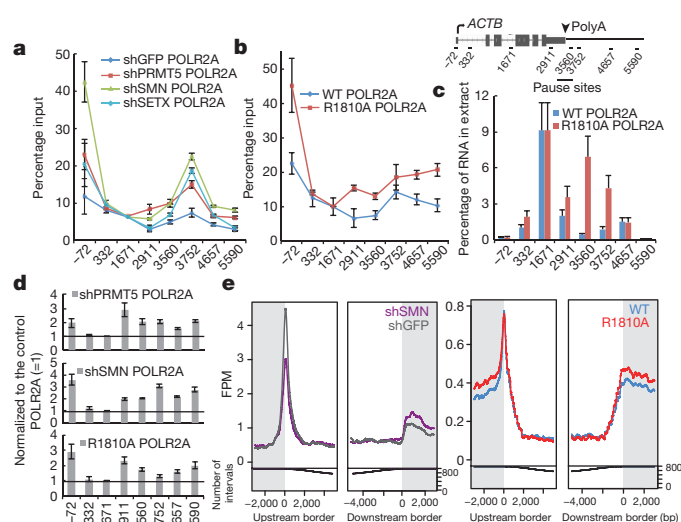
### SMN interacts with senataxin

To further understand the relationships of R1810me2s and SMN to transcription, we noted that SMN is known to interact with senataxin<sup>18</sup>, a DNA–RNA helicase encoded by the *SETX* gene that is important for termination by RNAP II<sup>18,19</sup>. We confirmed that SMN indeed interacts with senataxin (Fig. 1a, 3a, 3b); the interaction is apparently not mediated by RNAP II, as it persists after 3 days of treatment of cells with  $\alpha$ -amanitin to eliminate the bulk of the RNAP II (Fig. 3b). Moreover, the SMN–senataxin interaction is reduced when PRMT5 is stably knocked down, indicating that the interaction is probably mediated by the Tudor domain in SMN and an Arg-me2s modification on senataxin (Fig. 3c). Senataxin also co-precipitated with HA-tagged,  $\alpha$ -amanitin-resistant, wild-type RNAP II, and the interaction was reduced by the R1810A mutation (Fig. 2c), indicating that the senataxin–RNAP II interaction is likely stabilized by SMN. Consistent with this, in cells with stable shRNA-mediated or transient siRNA-mediated knockdown of PRMT5 or SMN, the association of senataxin with RNAP II was reduced (Fig. 3d, e and Extended Data Figs 1h–j and 3c).

Therefore, we used ChIP to test whether SMN and R1810 are important for senataxin recruitment to elongating transcription complexes. Senataxin ChIP in Raji cells depleted of endogenous POLR2A and expressing wild-type or R1810A mutant POLR2A revealed that R1810 is important for senataxin recruitment throughout the *ACTB* and *GAPDH* genes, including their termination regions (Fig. 3g and Extended Data Fig. 4d, e). SMN or PRMT5 knockdown revealed that both are important for senataxin recruitment (Fig. 3h and Extended Data Fig. 4f). Knockdown of senataxin itself showed that the senataxin antibody used for ChIP was specific for senataxin (Fig. 3h). Because senataxin is a termination factor<sup>18,19</sup>, these experiments showed that POLR2A (R1810), SMN, and PRMT5 should also be important for termination by RNAP II. Consistent with this, XRN2, a 5'→3' exonuclease involved in termination by RNAP II<sup>20</sup>, also co-precipitated with SMN and RNAP II (Fig. 1a).

### SMN and R1810me2s regulate termination

To investigate whether SMN and R1810me2s are important for R-loop resolution and termination, we carried out ChIP on  $\alpha$ -amanitin-resistant wild-type or (R1810A) mutant POLR2A after depleting endogenous  $\alpha$ -amanitin-sensitive POLR2A. There was enrichment of the R1810A mutant over wild-type RNAP II downstream of the cleavage and polyadenylation sites where RNAP II pauses and terminates transcription on *ACTB* (Fig. 4b, d and Extended Data Fig. 5) and *GAPDH* (Extended Data Fig. 6c, d), as detected by various anti-POLR2A monoclonal antibodies (4H8, H224, 8WG16). The effect of R1810 on RNAP II accumulation in termination regions was confirmed by carrying out a variant of the global run-on procedure in which nuclei are incubated with BrUTP and short run-on RNAs are isolated by binding to anti-BrU antibodies<sup>21</sup>. This experiment also indicated that the R1810A mutation leads to over-accumulation of active RNAP II downstream of the poly(A) sites

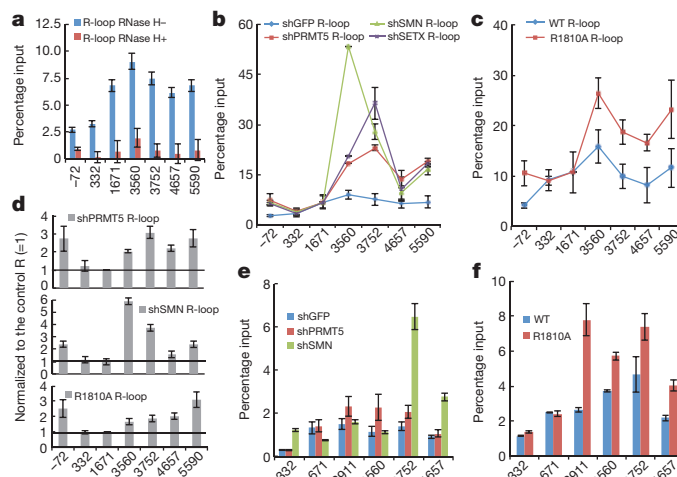


**Figure 4 | SMN and R1810me2s regulate transcription termination by RNAP II.** **a–d**, Quantification of RNAP II qPCR data from ChIP experiments using POLR2A antibodies (4H8, 8WG16, N20) (**a**, **b**, **d**) or nuclear run-on (**c**) in HEK293 (**a**) or Raji (**b**, **c**) cells, using the indicated *ACTB* primer positions, after stably knocking down (sh) PRMT5, SMN, or senataxin (**a**), or mutating R1810 to alanine (**b**). Error bars denote s.e.m. ( $n = 5$  biological replicates for **a**, **b**, **d**;  $n = 3$  biological replicates for **c**). **e**, RNAP II ChIP-seq in Raji cells (average density of ChIP fragments per million in the library for the 5% most highly expressed genes) for SMN (purple) versus GFP (grey) stable knockdowns (sh), or R1810A (red) versus wild-type (blue) POLR2A.

on *ACTB* and *GAPDH* (Fig. 4c and Extended Data Fig. 6e). When senataxin is stably knocked down, RNAP II similarly over-accumulates in the termination region of *ACTB*<sup>19</sup> (Fig. 4a), suggesting that senataxin and POLR2A R1810 are both needed for termination and release of RNAP II from DNA at the pause sites. Consistent with the idea that it is the lack of R1810me2s modification that leads to a failure of RNAP II release, stable shRNA-mediated or transient siRNA-mediated knock-down or CRISPR-mediated knockout of *SMN* or *PRMT5* led to similar over-accumulation of RNAP II in termination regions (Fig. 4a, d and Extended Data Figs 6b, d, 7 and 8a). We then performed ChIP-seq of RNAP II for shSMN versus shGFP, and POLR2A (R1810A) versus wild-type POLR2A samples. When compared to the control, SMN knock-down and the R1810A mutation on the CTD both led to a genome-wide RNAP II accumulation in termination regions of active genes (Fig. 4e and Extended Data Fig. 9).

Formation of R-loops by elongating RNAP II over pause sites downstream of poly(A) signals and their resolution by the senataxin helicase are important for recruiting the 5'→3' exonuclease XRN2 and termination by RNAP II<sup>19</sup>. The monoclonal antibody S9.6 binds RNA–DNA hybrids (Fig. 5a)<sup>19</sup>. When we used this antibody for DNA immunoprecipitation (DIP), we found that, like depletion of senataxin (Fig. 5b), mutation of R1810 (Fig. 5c, d) or knockdown of SMN or PRMT5 (Fig. 5b, d) led to an over-accumulation of R-loops in the termination region on the *ACTB* gene. A second method for R-loop detection employing a GFP fusion construct that includes the RNase H1 R-loop-binding domain (GFP–HB) was also used<sup>22</sup>. Between control and SMN CRISPR-mediated knockout cells that stably express GFP–HB, increased R-loop formation was detected in the termination regions in SMN knockout cells by GFP ChIP (Extended Data Fig. 8b). These experiments indicated that R1810me2s, PRMT5, and SMN lie upstream of senataxin in a common pathway important for R-loop resolution and transcription termination.

To determine whether the phenotypes for RNAP II and R-loop accumulation in the termination region may be relevant to the spinal muscular atrophy (SMA) disease state, fibroblast and B-lymphocyte cell lines were obtained for two children with SMA and their unaffected parents (Extended Data Fig. 8d). RNAP II ChIP and R-loop



**Figure 5 | SMN and R1810me2s are important for resolving R-loops created by elongating RNAP II and preventing DNA damage.**

**a–d**, Quantified DNA immunoprecipitation along *ACTB* using S9.6 antibody in HEK293 cells (**a**, **b**), with stable knock-down (short hairpin RNA, sh) of SMN, PRMT5 or senataxin (**a**), or treatment with RNase H (**a**), or in Raji cells after mutating R1810 to alanine (**c**). Error bars denote s.e.m. ( $n = 3$  to 5 biological replicates). **e**, **f**, Quantified *ACTB* gene qPCR data from  $\gamma$ H2AX ChIP experiments in HEK293 (**e**) or Raji (**f**) cells, after knocking down (sh) PRMT5 or SMN (**e**) or mutating R1810 to alanine (**f**). Error bars denote s.e.m. ( $n = 4$  biological replicates).

DIP showed that, compared to the control (average of the two parents), RNAP II and RNase H-sensitive R-loops tend to accumulate in the termination region of the *ACTB* gene in the SMA cells (Extended Data Fig. 8e, f).

Next, we examined why R-loop accumulation in termination regions might contribute to the neurodegeneration characteristic of SMA. By using RNA-seq to examine the effects of the POLR2A R1810A mutation, we confirmed that there are few gene expression changes aside from the previously noted upregulation of various snoRNAs and snRNAs<sup>7</sup> (data not shown). The GEMIN-containing SMN complex assembles various snRNPs that participate in spliceosome assembly<sup>23,24</sup>, and splicing deficiencies may account for SMA in a mouse model<sup>25–27</sup>. Similarly, the R1810me2s–SMN–SETX pathway may prevent SMA, at least in part, by affecting splicing, and more extensive analysis of our RNA-seq data revealed that the R1810A mutation causes many splicing alterations (data not shown). However, it is not clear whether these splicing changes are caused by upregulation of snRNA genes, changes in RNAP II elongation kinetics through the R1810me2s–SMN–SETX pathway<sup>28</sup>, or by recruitment by SMN of TDP-43 (also known as TARDBP) and FUS, which are known to interact with SMN<sup>29–31</sup> and affect splicing<sup>32–35</sup>.

Another possibility is that genome instability caused by less efficient removal of R-loops<sup>36</sup> may contribute to neurodegeneration. To test whether 3'-end R-loop accumulation can lead to DNA damage, we used a ChIP to assay the effects of the POLR2A R1810A mutation and the depletion of SMN or PRMT5 on the presence of  $\gamma$ H2AX, which accumulates at sites of DNA damage<sup>37,38</sup> (Fig. 5e, f and Extended Data Fig. 10a, b). As expected, we observed accumulation of  $\gamma$ H2AX and increased  $\gamma$ H2AX:H2AX ratio in the *ACTB* gene termination region where RNAP II and R-loops accumulate. Therefore, accumulation of termination region R-loops may generally lead to DNA damage and genome instability, as is the case when senataxin is depleted<sup>37</sup>.

## Discussion

The finding that R1810 in the RNAP II CTD is symmetrically dimethylated by PRMT5 is substantiated by many observations described here. First, immunoprecipitated RNAP II is recognized by two different antibodies in a western blot, SYMM10 and Y12, specific for Arg-me2s, as well as antibodies raised against an R1810me2s peptide and specific for



the Arg-me2s modification. Second, the SDS gel mobility of this band is the same as that of the largest subunit of RNAP II. Third, recognition of the modification by the antibodies depends on R1810. Fourth, the modification recognized by the antibodies depends on PRMT5. Fifth, POLR2A R1810 and PRMT5 are needed for the association with RNAP II and the recruitment to transcribed regions of SMN, whose Tudor domain is specific for Arg-me2s. Symmetric dimethylation of R1810 in the POLR2A CTD causes the direct recruitment of SMN and indirect recruitment of the RNA–DNA helicase senataxin. This is followed by R-loop resolution and, in termination regions, recruitment of the 5'→3' exonuclease XRN2 and efficient termination by RNAP II (see model in Extended Data Fig. 10c). Mutation of R1810 or depletion of SMN or PRMT5 may also lead to RNAP II accumulation in promoter-proximal regions, perhaps because some RNAP II molecules that pause downstream of promoters<sup>21</sup> terminate prematurely in a process that depends on R1810me2s and SMN.

POLR2A R1810 is both symmetrically dimethylated by PRMT5 and asymmetrically dimethylated by CARM1 (ref. 7), but it is not uncommon for type I and type II PRMTs to compete for deposition of me2a and me2s, respectively, on the same substrates<sup>39</sup>. Other examples of alternative modifications of the same Arg residue are PRMT5 antagonizing PRMT1-dependent activation of the histone H4R3me2a mark by depositing repressive me2s marks on histones H4R3 and H3R8 (refs 16, 40, 41), and alternative modifications of R698 of the elongation factor SPT5 by PRMT1 and PRMT5 to regulate its role in RNAP II elongation<sup>42</sup>.

SMN, which recognizes R1810me2s, can self-aggregate through its N-terminal K-rich domain and C-terminal YG box, amplifying its potential for Arg-me2s-mediated protein–protein interactions<sup>24,43,44</sup>. Although the human genome contains two SMN genes (*SMN1*, *SMN2*), mutating one copy of *SMN1* decreases total SMN protein enough to cause SMA, with disease severity reflecting insufficiency of the remaining SMN for its full range of functions<sup>24</sup>.

SMN oligomers facilitate snRNP assembly by binding Arg-me2s modifications on Sm proteins<sup>23,24</sup>. Many RNAP II elongation and termination factors and proteins implicated in neurodegeneration also contain dimethylated Arg, including XRN2, three subunits of CPSF (CPSF1, CPSF5, and CPSF6), CSTF2, three poly(A)-binding proteins (PABP1, PABP 2, and PABP4), RBBP6, WDR33, PCF11, SPT5, CTDP1, DHX9, FUS, and EWSR1<sup>45–47</sup>. It is possible, therefore, that SMN may help assemble an R-loop resolving complex on the RNAP II CTD by binding several Arg-me2s-modified termination factors.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 6 January 2014; accepted 20 November 2015.**

**Published online 23 December 2015.**

- Burawski, S. Progression through the RNA polymerase II CTD cycle. *Mol. Cell* **36**, 541–546 (2009).
- Jonkers, I. & Lis, J. T. Getting up to speed with transcription elongation by RNA polymerase II. *Nature Rev. Mol. Cell Biol.* **16**, 167–177 (2015).
- Porrua, O. & Libri, D. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nature Rev. Mol. Cell Biol.* **16**, 190–202 (2015).
- Venkatesh, S. & Workman, J. L. Histone exchange, chromatin structure and the regulation of transcription. *Nature Rev. Mol. Cell Biol.* **16**, 178–189 (2015).
- Mayer, A. et al. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* **336**, 1723–1725 (2012).
- Hsin, J. P., Sheth, A. & Manley, J. L. RNAP II CTD phosphorylation on threonine-4 is required for histone mRNA 3' end processing. *Science* **334**, 683–686 (2011).
- Sims, R. J. III et al. The C-terminal domain of RNA polymerase II is modified by site-specific methylation. *Science* **332**, 99–103 (2011).
- Yang, Y. et al. Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. *Mol. Cell* **53**, 484–497 (2014).
- Yang, Y. et al. TDRD3 is an effector molecule for arginine-methylated histone marks. *Mol. Cell* **40**, 1016–1023 (2010).
- Yang, Y. et al. PRMT9 is a type II methyltransferase that methylates the splicing factor SAP145. *Nat. Commun.* **6**, 6428 (2015).

- Hadjikyriacou, A., Yang, Y., Espejo, A., Bedford, M. T. & Clarke, S. G. Unique features of human protein arginine methyltransferase 9 (PRMT9) and its substrate RNA splicing factor SF3B2. *J. Biol. Chem.* **290**, 16723–16743 (2015).
- Licciardo, P. et al. The FCP1 phosphatase interacts with RNA polymerase II and with MEP50 a component of the methylosome complex involved in the assembly of snRNP. *Nucleic Acids Res.* **31**, 999–1005 (2003).
- Antonyasmy, S. et al. Crystal structure of the human PRMT5:MEP50 complex. *Proc. Natl Acad. Sci. USA* **109**, 17960–17965 (2012).
- Friesen, W. J. et al. A novel WD repeat protein component of the methylosome binds Sm proteins. *J. Biol. Chem.* **277**, 8243–8247 (2002).
- Chen, C., Nott, T. J., Jin, J. & Pawson, T. Deciphering arginine methylation: Tudor tells the tale. *Nature Rev. Mol. Cell Biol.* **12**, 629–642 (2011).
- Bedford, M. T. & Clarke, S. G. Protein arginine methylation in mammals: who, what, and why. *Mol. Cell* **33**, 1–13 (2009).
- Sikorsky, T. et al. Recognition of asymmetrically dimethylated arginine by TDRD3. *Nucleic Acids Res.* **40**, 11748–11755 (2012).
- Suraweera, A. et al. Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. *Hum. Mol. Genet.* **18**, 3384–3396 (2009).
- Skourti-Stathaki, K., Proudfoot, N. J. & Gromak, N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* **42**, 794–805 (2011).
- West, S., Gromak, N. & Proudfoot, N. J. Human 5'→3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432**, 522–525 (2004).
- Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).
- Bhatia, V. et al. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* **511**, 362–365 (2014).
- Battle, D. J. et al. The SMN complex: an assembly machine for RNPs. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 313–320 (2006).
- Burghes, A. H. & Beattie, C. E. Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick? *Nature Rev. Neurosci.* **10**, 597–609 (2009).
- Zhang, Z. et al. SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. *Cell* **133**, 585–600 (2008).
- Bäumer, D. et al. Alternative splicing events are a late feature of pathology in a mouse model of spinal muscular atrophy. *PLoS Genet.* **5**, e1000773 (2009).
- Zhang, Z. et al. Dysregulation of synaptogenesis genes antecedes motor neuron pathology in spinal muscular atrophy. *Proc. Natl Acad. Sci. USA* **110**, 19348–19353 (2013).
- Muñoz, M. J. et al. DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. *Cell* **137**, 708–720 (2009).
- Ling, S. C. et al. ALS-associated mutations in TDP-43 increase its stability and promote TDP-43 complexes with FUS/TLS. *Proc. Natl Acad. Sci. USA* **107**, 13318–13323 (2010).
- Kim, S. H., Shanware, N. P., Bowler, M. J. & Tibbetts, R. S. Amyotrophic lateral sclerosis-associated proteins TDP-43 and FUS/TLS function in a common biochemical complex to co-regulate HDAC6 mRNA. *J. Biol. Chem.* **285**, 34097–34105 (2010).
- Yamazaki, T. et al. FUS–SMN protein interactions link the motor neuron diseases ALS and SMA. *Cell Rep.* **2**, 799–806 (2012).
- Lagier-Tourenne, C. et al. Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nature Neurosci.* **15**, 1488–1497 (2012).
- Polymenidou, M. et al. Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nature Neurosci.* **14**, 459–468 (2011).
- Ishigaki, S. et al. Position-dependent FUS–RNA interactions regulate alternative splicing events and transcriptions. *Sci. Rep.* **2**, 529 (2012).
- Rogelj, B. et al. Widespread binding of FUS along nascent RNA regulates alternative splicing in the brain. *Sci. Rep.* **2**, 603 (2012).
- Aguilera, A. & Garcia-Muse, T. R loops: from transcription byproducts to threats to genome stability. *Mol. Cell* **46**, 115–124 (2012).
- Hatchi, E. et al. BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Mol. Cell* **57**, 636–647 (2015).
- Sollier, J. et al. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol. Cell* **56**, 777–785 (2014).
- Dhar, S. et al. Loss of the major type I arginine methyltransferase PRMT1 causes substrate scavenging by other PRMTs. *Sci. Rep.* **3**, 1311 (2013).
- Pal, S., Vishwanath, S. N., Erdjument-Bromage, H., Tempst, P. & Sif, S. Human SWI/SNF-associated PRMT5 methylates histone H3 arginine 8 and negatively regulates expression of ST7 and NM23 tumor suppressor genes. *Mol. Cell Biol.* **24**, 9630–9645 (2004).
- Pal, S. & Sif, S. Interplay between chromatin remodelers and protein arginine methyltransferases. *J. Cell. Physiol.* **213**, 306–315 (2007).
- Kwak, Y. T. et al. Methylation of SPT5 regulates its interaction with RNA polymerase II and transcriptional elongation properties. *Mol. Cell* **11**, 1055–1066 (2003).
- Talbot, K. et al. Missense mutation clustering in the survival motor neuron gene: a role for a conserved tyrosine and glycine rich region of the protein in RNA metabolism? *Hum. Mol. Genet.* **6**, 497–500 (1997).



44. Young, P. J. *et al.* The exon 2b region of the spinal muscular atrophy protein, SMN, is involved in self-association and SIP1 binding. *Hum. Mol. Genet.* **9**, 2869–2877 (2000).
45. Boisvert, F. M., Chenard, C. A. & Richard, S. Protein interfaces in signaling regulated by arginine methylation. *Sci. STKE* **2005**, re2 (2005).
46. Boisvert, F. M., Cote, J., Boulanger, M. C. & Richard, S. A proteomic analysis of arginine-methylated protein complexes. *Mol. Cell. Proteomics* **2**, 1319–1330 (2003).
47. Uhlmann, T. *et al.* A method for large-scale identification of protein arginine methylation. *Mol. Cell. Proteomics* **11**, 1489–1499 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank J. Manley for GST–CTD constructs; D. Reinberg for RNAP II me2a antibodies; D. Eick for wild-type and RNAP II (R1810A) constructs; and S. Leppla for purified S9.6 antibodies. We also thank D. Torti and D. Leung for Illumina library preparation and sequencing, T. Hajian for the purified PRMT5–WDR77 complex, J. Li for purified 8WG16 antibodies, and D. Durocher for constructive criticism and advice during the course of this work. This project was supported by the Ontario Research Fund from the Ontario Ministry of Research and Innovation (to J.F.G. and T.P.) and by CIHR Operating

Grants to J.F.G. and B.J.B. D.Y.Z. was supported by a National Science and Engineering Research Council of Canada Studentship and an Ontario Graduate Scholarship. U.B. was supported by a long-term Postdoctoral Fellowship from HFSP. B.J.B. holds the University of Toronto Banbury Chair in Medical Research.

**Author Contributions** J.F.G. supervised the project. D.Y.Z. performed the experiments. J.F.G. and D.Y.Z. wrote the manuscript. T.J.P., B.J.B., Z.N., and G.G. commented on experiments and edited the manuscript. G.G. prepared the FITC peptides. F.W.S. performed ChIP-seq experiments. U.B. and Y.L. performed computational data analysis for ChIP-seq. Vectors for shRNAs were provided by J.M., and G.Z. generated stable shRNA-mediated knockdown cell lines. Z.N. generated CRISPR knockout cell lines. J.M. and K.L. provided the purified Tudor domains. M.V. provided the purified PRMT5–WDR77 complex. W.L. performed isothermal titration calorimetry assays.

**Author Information** Data for ChIP-seq analyses have been deposited at GEO with the accession code GSE73379. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.F.G. ([jack.greenblatt@utoronto.ca](mailto:jack.greenblatt@utoronto.ca)).

## METHODS

The number of times each of the experiments described in this work was performed and the raw western blots used in the figures are shown in the Supplementary Information file. The investigators were not blinded to allocation during experiments and outcome assessment.

**In vitro PRMT5 methyltransferase assays.** *In vitro* methylation assays (30 µl) containing 100 µM of biotinylated 13-mer CTD peptides centred around R1810 or R1603, 0.1–2 µM PRMT5–WDR77 complex, 100 µM tritiated SAM (PerkinElmer, catalogue number NET155V250UC), 20 mM Tris-HCl pH8, 0.01% TritonX-100 (Sigma catalogue number T8532), and 5 mM DTT were incubated at room temperature for 2 h. The biotinylated peptides were then precipitated with 20 µl streptavidin-agarose (Invitrogen catalogue number SA100), washed with buffer (above), mixed with 5 ml Scintivase BD Cocktail (Fisher Chemical, catalogue number SX18-4), and counted using a Beckman Liquid Scintillation Counter (LS 6500). Other similar reactions contained recombinant GST proteins (GST alone, GST-N-CTD or GST-C-CTD), 0.1–2 µM PRMT5–WDR77 and 100 µM tritiated SAM. These reactions were precipitated with glutathione beads (Invitrogen catalogue number G2879), and the beads were washed and eluted with 500 µl of 20 mM L-glutathione (pH 8) for counting.

**Peptides, GST recombinant proteins, recombinant Tudor domains, fluorescence polarization and isothermal titration calorimetry assays.** FITC- and biotin-labelled CTD peptides containing R1603 (SPAYEPRSPGGYT) and R1810 (YSPSPRYTPQSP) were prepared on a Prelude peptide synthesizer (Protein Technologies, Tucson Arizona) using Fmoc (9-fluorenyl methoxycarbonyl) solid-phase chemistry. Dimethyl arginine derivatives were prepared using Fmoc-SDMA(Boc)2-ONa or Fmoc-ADMA (Pbf)-OH reagents (Novabiochem, Germany). Peptides were purified using C18 reverse-phase HPLC and authenticated using mass spectrometry.

Constructs for GST recombinant protein expression (GST-N-CTD: contains repeats 1–29; or GST-C-CTD: contains repeats 24–52) were expressed in BL21 bacteria and purified following the standard glutathione bead purification protocol. Bacterial expression constructs and purified Tudor domains from TDRD3, SMN, SPF30, TDRD1, TDRD2, TDRD9 and TDRD11 (also known as SND1) were described previously<sup>48</sup>.

Fluorescence polarization assays were carried out as described before<sup>48</sup>. The buffer used in the fluorescence polarization assay was 20 mM Tris pH 7.5, 50 mM NaCl, 1 mM DTT and 0.01% Triton X-100. An excitation wavelength of 485 nm and an emission wavelength of 528 nm were used. The data were obtained at 25 °C and corrected by subtracting the label-free peptide background. The data were collected by the Synergy 2 (BioTec, USA) fluorescence polarization program and were fitted to a one-site binding model using Origin 7 (MicroCal). The  $K_d$  values are from the average of three measurements.

For isothermal titration calorimetry, the concentrated protein was diluted in 20 mM Tris, pH 7.5, 150 mM NaCl. The lyophilized peptides were dissolved in the same buffer and pH was adjusted by adding NaOH. Peptide concentrations were estimated from the molecular weight. All the measurements were performed at 25 °C, using a VP-ITC microcalorimeter. Protein with a concentration of 50 µM was placed in the cell chamber, and the peptides with a concentration of 1 mM in syringe were injected in 25 successive injections with a spacing of 180 s and a reference power of 13 mcal s<sup>-1</sup>. Data were fitted using the single-site binding model within the Origin software package (MicroCal).

**Cell culture, shRNA knockdowns, siRNA knockdowns, CRISPR knockouts, and electroporation.** There was no evidence of mycoplasma contamination of the cell lines used in this work as judged by staining of fixed cells with DAPI. Raji cells were cultured in RPMI (SLRI media facility) plus 10% FBS (Sigma catalogue number F1051) and 1% glutamate, and stably transduced cells were maintained with 500 µg ml<sup>-1</sup> G418 (Gibco catalogue number 11811031). HEK293 cells were grown in DMEM (SLRI media facility) plus FBS (Sigma catalogue number F1051), and stably transduced cell lines were maintained with 2 µg ml<sup>-1</sup> puromycin (Sigma catalogue number p8833). shRNAs in lentivirus vectors were used to stably transduce cell lines using an established protocol<sup>49</sup>. siRNA knockdowns for HEK293 cells were performed with 50 nM SMARTpool siRNAs with PepMute siRNA transfection reagent (SigmaGen Laboratory catalogue number SL100566) for 3 days. SMARTpool On-Target plus siRNAs against human PRMT5 (catalogue number L-015817) and SMN (catalogue number L-011108) were purchased from Thermo Scientific.

For CRISPR-mediated gene knockouts, CRISPR/Cas9 plasmids (pCMV-Cas9-GFP) were purchased from Sigma-Aldrich which express scrambled guide RNA, or guide RNA that targets the SMN1 gene. 2 µg of the plasmids were transfected into HEK293 cells, and 1 day after transfection, cells were sorted by BD FACSaria flow cytometry (Donnelly Centre, University of Toronto) and single GFP-positive cells were plated into a 48-well plate. The expression levels of SMN in each clone were detected by western blotting.

Raji cells with stable expression of HA-tagged wild type or POLR2A (R1810A) constructs were generated by electroporation (10 µg of plasmid DNA per 10<sup>7</sup> cells), followed by selection and maintenance with G418 (0.5 mg ml<sup>-1</sup>). α-amanitin treatment was carried out with 2 µg ml<sup>-1</sup> α-amanitin for 3 days for Co-IP and ChIP experiments involving HA-tagged wild-type or POLR2A (R1810A). The transfection of the GFP-HB transgene for R-loop detection into HEK293 cell lines was performed with the FuGENE Transfection reagent (Roche, catalogue number E269A).

**Disease cell lines.** SMA disease relevant and control fibroblast and B-lymphocyte cell lines were obtained from the Coriell Institute (Family 553: GM03813, GM03814, GM03815; Family 3042: GM23686, GM23687, GM23688), and were grown in conditions as instructed by the Coriell Institute. The cells were collected and fixed for RNAP II ChIP and R-loop DIP.

**Immunoprecipitation (IP) and western blots.** IP was performed with RIPA buffer (140 mM NaCl, 10 mM Tris pH 7.6–8.0, 1% Triton, 0.1% sodium deoxycholate, 1 mM EDTA) containing protease inhibitors (Roche catalogue number 05892791001) and benzonase (Sigma E1014). 10<sup>7</sup> to 2 × 10<sup>7</sup> cells were lysed on ice for 25 min by vortexing and forcing them through a 27 gauge needle. After centrifuging at 13,000 r.p.m. for 10 min at 4 °C, the supernatant was incubated with 25 µl (1:10 dilution) of protein G beads (Invitrogen catalogue numbers 10-1243 and 10003D) and 1–2 µg of antibodies for 4 h to overnight. The samples were washed 3 times with RIPA buffer and boiled in SDS gel sample buffer. To detect R1810me2s or R1810me2a modifications on POLR2A, alkaline phosphatase (Roche catalogue number 10108138001) treatment (5 µl) at 37 °C for 30 min was performed for POLR2A immunoprecipitated samples before boiling. Samples were run using 7.5–10% SDS–PAGE and transferred to PVDF membranes (Bio-Rad catalogue number 162-0177) using a trans-blot semi-dry electrophoretic transfer Cell (BioRad catalogue number 170-3940). Primary antibodies were used at 1:250 to 1:1,000 dilutions for incubation overnight, and horseradish peroxidase-conjugated goat anti-mouse, anti-rabbit, or anti-rat secondary antibodies were used at 1:10,000 (Dako catalogue number P0450). Blots were developed using SuperSignal West Pico or Femto (Thermo Scientific catalogue numbers 34079 and 34094). Blots were quantified using ImageJ software.

A Hoefer slot blot system (Fisher Scientific catalogue number 11509543) was used to assay R1810me2s antibody specificities following the manufacturer's protocol.

**Chromatin immunoprecipitation (ChIP) and DNA immunoprecipitation (DIP).** ChIP was performed using the EZ-ChIP A chromatin immunoprecipitation kit (Millipore catalogue number 17-371) or similar homemade solutions according to the manufacturer's instructions. Antibodies were used in the 1–2 µg range, and IgG was used as a background control. DIP was performed according to ref. 50 with minor modifications. DIP was performed following the ChIP protocol except that, after the nuclear lysis and sonication, genomic DNA was de-crosslinked in ChIP elution buffer containing 5 M NaCl at 65 °C overnight. DNA was purified with the Qiaex II kit (Qiagen catalogue number 20021) for PCR product purification and eluted in water. DIP was carried out overnight with 25 µl of Dynabeads protein G beads (Invitrogen catalogue number 100-03D) and 1 µg of antibody purified from the S9.6 hybridoma cell line<sup>51</sup> that recognizes RNA–DNA hybrids. Immunoprecipitated and input DNAs were used as templates for qPCR. DIP RNase-sensitivity analysis was carried out by adding 50 U of RNase H (Invitrogen catalogue number 18021-014) in 10 × RNase H buffer (75 mM KCl, 50 mM Tris pH 8.3, 3 mM MgCl<sub>2</sub>, 10 mM DTT) with 4% glycerol and 20 µg ml<sup>-1</sup> BSA before immunoprecipitation. The RNase H treatment was performed for 2 h at 37 °C.

For comparing POLR2A ChIP and S9.6 DIP signals on the ACTB gene, wild-type or control knockdown signals were normalized to 1, and the R1810A mutant or knockdown samples were adjusted such that the ratio for the intron 3 (1671) position was set to 1. Similarly, for the GAPDH gene, the ratio for the intron 5 (2436) position was set to 1. ChIP data for senataxin and SMN were expressed as ratios to the ChIP data for POLR2A. Error bars represent biological replicates, except where indicated otherwise.

**Nuclear run-ons (NRO).** NRO was performed according to Skourti-Stathaki *et al.* with modifications<sup>19,21</sup>. Approximately 10<sup>7</sup> cells were incubated on ice in swelling buffer (10 mM Tris-Cl pH 7.5, 2 mM MgCl<sub>2</sub>, 3 mM CaCl<sub>2</sub>) for 5 min, and were pelleted. Pellets were resuspended in 1 ml lysis buffer (swelling buffer containing 0.5% NP40, 10% glycerol, and 2 U ml<sup>-1</sup> RNaseOUT (Invitrogen catalogue number 10777-019)) and pipetted for lysis, followed by centrifugation. The pellet was resuspended in 1 ml freezing buffer (50 mM Tris-Cl pH 8.3, 40% glycerol, 5 mM MgCl<sub>2</sub>, 0.1 mM EDTA). Reactions contained 100 µl resuspended nuclei, 100 µl reaction buffer (40 mM Tris pH 7.9, 300 mM KCl, 10 mM MgCl<sub>2</sub>, 40% glycerol, 2 mM DTT), 500 µM rNTPs (ATP, CTP, GTP) (GE catalogue number 27-2025-01), including 125 µM UTP as a negative control or Br-UTP (Invitrogen catalogue number B21551) for 30 min at 30 °C. 3 µl BrdU antibody (Sigma catalogue number B8434) was pre-conjugated to 30 µl Dynabeads protein G beads with 10 µg tRNA (Invitrogen catalogue number 15401) as block in 100-RSB buffer (10 mM Tris

pH 7.4, 100 mM NaCl, 2.5 mM MgCl<sub>2</sub>, 0.4% Triton X-100) for 2 h at 4°C. The RNA was extracted using TRIzol (Invitrogen catalogue number 15596-026) and was heat fragmented at 95°C for 8 min. RNA was then mixed with beads and BrdU antibody for 2 h at 4°C in 500 µl 100-RSB buffer, 100 U ml<sup>-1</sup> RNase OUT, 400 U ml<sup>-1</sup> DNase I (Invitrogen catalogue number 18047-019). Immunoprecipitated RNA was washed three times with 100-RSB buffer.

**Primer information.** Primers used in ChIP, DIP, and NRO are listed here<sup>19,52</sup>.

For the *ACTB* gene: -72.fw CCGAAAGTTGCCTTTATGGC, -72.rev CAAAGCGGAGGCTCTGTGC; 332.fw CGGGGTCTTTGTCTGAGC, 332.rev CAGTTAGCGCCCAAAGGAC; 1671.fw TAACACTGGCTCGTGTGACAA, 1671.rev AAGTGCAAAGAACACGGCTAA; 2661.fw GGAGCTGTACATCCAGGGTC, 2661.rev TGCCACTGGCTCGTGTGACAA; 2911.fw TGCGCA GAAAACAAGATGAG, 2911.rev GTCACCTTCACCGTTCCAGT; 3560.fw TTACCCAGAGTGCAGGTGTG, 3560.rev CCCCATAAGCAGGAACAGA; 3752.fw GGGACTATTTGGGGGTGTCT, 3752.rev TCCCATAGGTGAA GGCAAAG; 4657.fw TGGGCCACTTAATCATTCAAC, 4657.rev CCTCACTTC CAGACTGACAGC; 5590.fw CAGTGGTGTGGTGTGATCTTG, 5590.rev GGC AAAACCTGTATCTGTGA.

For the *GAPDH* gene: 55.fw CTCCTGTTTCGACAGTCAGC, 55.rev TTCAGGCGCTCCCTAGC; 1407.fw CACCCTGGTCTGAGGTTAAATATAG, 1407.rev GTGGGAGCACAGGTAAGT; 2436.fw ATAGGCGAGATCCCTCCAA, 2436.rev TGAAGACGCCAGTGGAC; 3882.fw CCCTGTGCTCAACCACT, 3882.rev CTCACCTTGACACAAGCC; 4511.fw AGATGTGTCAAGG TGACTTAT, 4511.rev TAGGTCCCAGTACACGC; 5196.fw GTCTCAGTGTAT GACAGACACG, 5196.rev TGTATGTGCGCTCAGGG.

**Chromatin immunoprecipitation and sequencing analysis (ChIP-seq).** Chromatin immunoprecipitation was performed as before<sup>53</sup>. In brief, 10<sup>7</sup> to 10<sup>8</sup> cells were crosslinked for 10 min in 1% formaldehyde. Lysates were sonicated to a DNA fragment length range of 200–300 bp using a Bioruptor (Diagenode). RNAP II was immunoprecipitated with 2 µg of antibodies and Dynabeads Protein G (Invitrogen). Subsequently, crosslinks were reversed at 65°C overnight and bound DNA fragments were purified (EZ-10 spin column PCR product purification kit, Bio Basic). Sequencing libraries were constructed using the TruSeq ChIP sample prep kit (Illumina) according to the manufacturer's instructions. Libraries were sequenced (single-end reads) on the Illumina HiSeq 2500 to a minimum depth of 30 million reads, obtaining at least 10 million unique reads per sample.

ChIP-seq analysis was performed chiefly as before<sup>54</sup>. For ChIP-seq, reads in FASTQ format were mapped to the human genome (hg19) using Bowtie 2 (ref. 55) with local alignment, duplicate reads were removed, and reads were extended to 300 bp. Pileups—the number of fragments overlapping each genomic bp—were calculated, and were normalized by million mappable reads in the ChIP-seq library. Normalized pileups from different replicates were then averaged to create FPM (fragments per million reads). Data for ChIP-seq analyses have been deposited in GEO with the accession code GSE73379.

**RNA extraction and qPCR.** RNA was used for cDNA synthesis with the SuperScript VILO Kit (Invitrogen catalogue number 11754). PCR was performed using the Phusion-high fidelity PCR kit (Thermo Scientific catalogue number F-553S), and qPCR was performed with Fast SYBR Green Master qPCR mix, using the Applied Biosystems 7300 real time PCR System (catalogue number 4406984). qPCR consisted of 40 cycles of 95°C for 15 s and 55°C for 30 s, and a final cycle (95°C for 15 s and then 60°C) generated a dissociation curve. Input DNA or RNA reverse transcribed into cDNA were used to calculate the per cent enrichment in the immunoprecipitated samples.

**Antibodies, constructs and reagents.** Anti-CTD R1810me2s antibody was raised in rabbits using a KLH-conjugated CTD peptide from POLR2A (amino acids 1806–1813) that carried an R1810me2s modification. KLH conjugation was

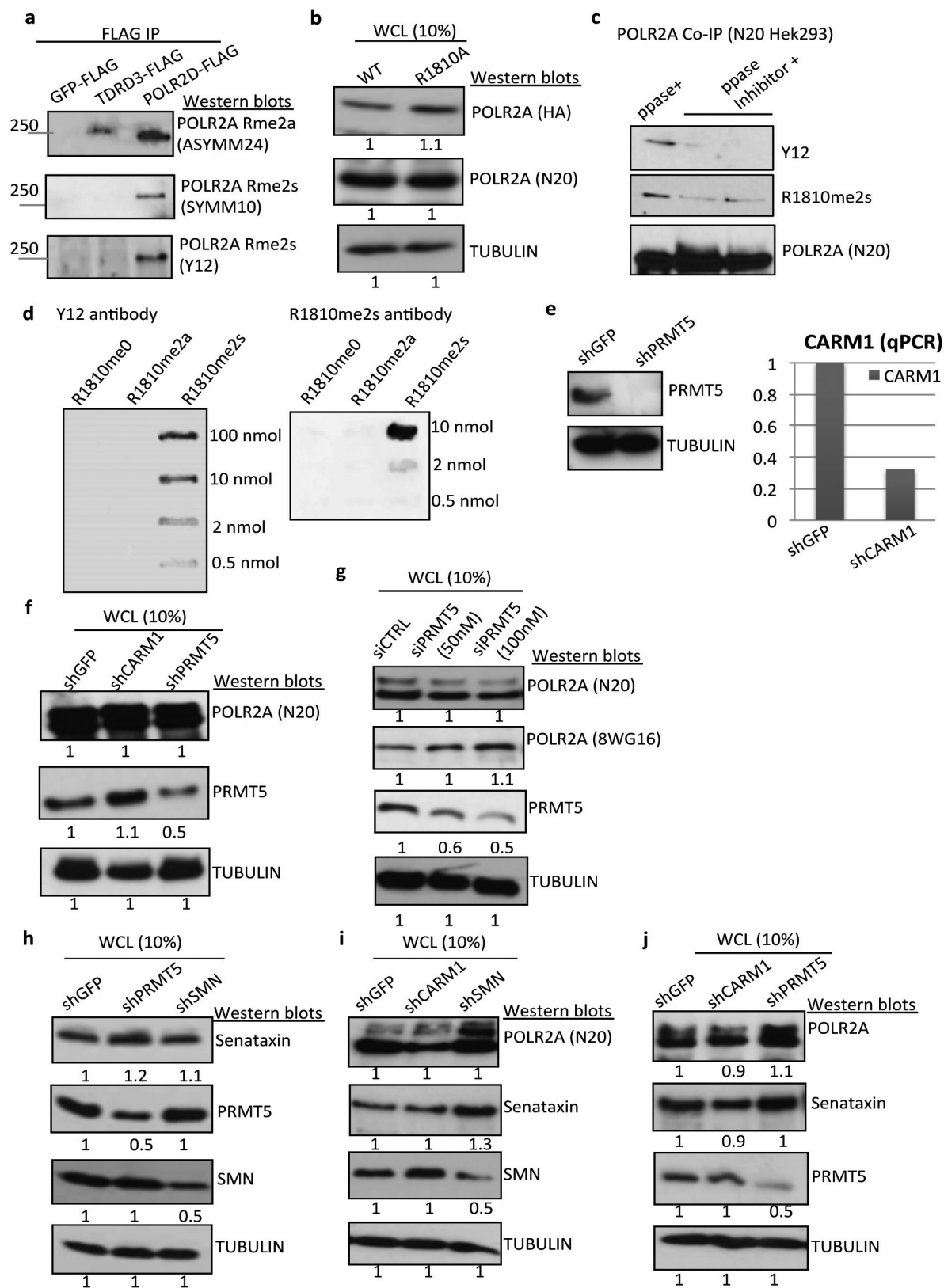
performed using an N-terminal cysteine residue (Cedarlane). R1810me2s-specific antibodies were enriched by flowing the serum through a column containing an R1810me0 peptide conjugated to SulfoLink Coupling Resin (Thermo Scientific catalogue number 20401).

GST fusion constructs containing CTD N-terminal repeats 1–29 and C-terminal repeats 33–52 were provided by J. Manley<sup>56</sup>. Flp-in T-REX GFP-HB fusion construct that contains the R-loop binding domain of RNase H was provided by A. Aguilera<sup>52</sup>. The ORFs for POLR2D, TDRD3, PRMT5, and SMN came from the plasmid collection at Harvard. CMV promoter-driven Flag-tagged TDRD3, GFP, and PRMT5 constructs for HEK293 cell culture were generated using the MAPLE system as previously described<sup>49</sup>.  $\alpha$ -amanitin-resistant wild-type and R1810A mutant POLR2A constructs were kindly provided by D. Eick<sup>7</sup>. RNAP II R1810 me2a antibody was provided by D. Reinberg<sup>7</sup>. We obtained the POLR2A pSer2 and pSer5 antibodies from the Eick laboratory (S2P: 3E10; S5P: 3E8). We obtained the S9.6 antibody for R-loop DIP from S. Lepplä. 8WG16 monoclonal antibody against unphosphorylated CTD repeats of POLR2A was prepared in our laboratory.

Commercial antibodies were as follows: Y12 (Abcam monoclonal antibody catalogue number ab3138); Sym10 (Millipore polyclonal antibody catalogue number 07-412); Asym24 (Millipore polyclonal antibody catalogue number 07-414); TDRD3 (Santa Cruz polyclonal antibody catalogue number C-20); HA (Sigma monoclonal antibody catalogue number H9658); PRMT5 (Upstate polyclonal antibody catalogue number C7-405, Santa Cruz monoclonal antibody catalogue number sc-22132); SMN (Santa Cruz polyclonal antibody catalogue number H-195); SETX (for ChIP and IP (Novus Biologicals polyclonal antibody catalogue number NB100-57543) and for western blots (Bethyl Lab polyclonal antibody catalogue number A301-104A)); XRN2 (Santa Cruz polyclonal antibody catalogue number sc-99237); POLR2A N20 (Santa Cruz polyclonal antibody catalogue number sc-899); POLR2A 4H8 (Abcam monoclonal antibody catalogue number ab5408); POLR2A H224 (Santa Cruz polyclonal antibody catalogue number sc9001);  $\gamma$ H2AX (Millipore, catalogue number 05-636); H2AX (Millipore, catalogue number 07-627); tubulin (Sigma monoclonal antibody catalogue number T8328); Flag (Sigma monoclonal antibody catalogue number F1804); GFP (Abcam polyclonal antibody catalogue number 290); IgG negative controls for ChIP and IP (Millipore polyclonal antibody catalogue number 12-370).  $\alpha$ -amanitin was purchased from Sigma (catalogue number 23109-05-9).

48. Liu, K. *et al.* Crystal structure of TDRD3 and methyl-arginine binding characterization of TDRD3, SMN and SPF30. *PLoS ONE* **7**, e30375 (2012).
49. Mak, A. B. *et al.* A lentiviral functional proteomics approach identifies chromatin remodeling complexes important for the induction of pluripotency. *Mol. Cell. Proteomics* **9**, 811–823 (2010).
50. El Hage, A., French, S. L., Beyer, A. L. & Tollervey, D. Loss of topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev.* **24**, 1546–1558 (2010).
51. Boguslawski, S. J. *et al.* Characterization of monoclonal antibody to DNA:RNA and its application to immunodetection of hybrids. *J. Immunol. Methods* **89**, 123–130 (1986).
52. Glover-Cutter, K., Kim, S., Espinosa, J. & Bentley, D. L. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nature Struct. Mol. Biol.* **15**, 71–78 (2008).
53. Schmidt, D. *et al.* ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods* **48**, 240–248 (2009).
54. Braunschweig, U. *et al.* Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* **24**, 1774–1786 (2014).
55. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).
56. Ryan, K., Murthy, K. G., Kaneko, S. & Manley, J. L. Requirements of the RNA polymerase II C-terminal domain for reconstituting pre-mRNA 3' cleavage. *Mol. Cell. Biol.* **22**, 1684–1692 (2002).

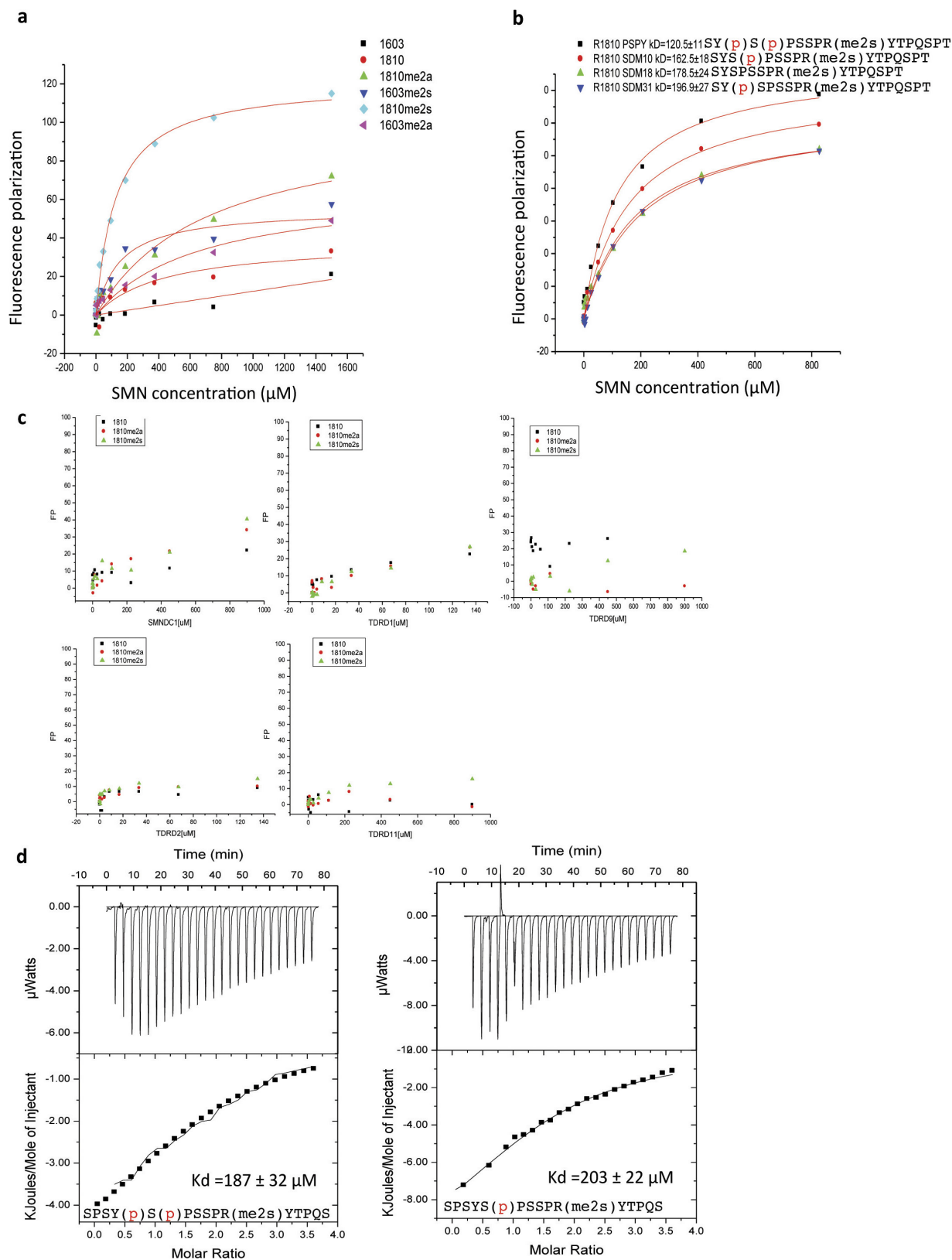




Extended Data Figure 1 | See next page for caption.

**Extended Data Figure 1 | The R1810me2s and R1810me2a modifications on POLR2A depend on PRMT5 and CARM1, respectively.** **a**, POLR2A carries Rme2a and Rme2s modifications. Whole-cell lysates (WCL) from HEK293 cells stably expressing Flag-tagged TDRD3 or the POLR2D subunit of RNAP II were used for immunoprecipitation using beads conjugated with M2 anti-Flag antibody, and the precipitates were western blotted with the indicated antibodies. Cells expressing Flag-GFP were used as a negative control. Precipitated TDRD3 and POLR2D contained POLR2A with the Arg-me2a modification (ASYMM24 antibody), whereas precipitated POLR2D, and not TDRD3, contained POLR2A with the Arg-me2s modification (SYMM10 and Y12 antibodies). **b**, Whole-cell lysate western blot controls for Figure 1b. **c**, Y12 and R1810me2s recognition of RNAP II

CTD R1810me2s is blocked by surrounding phosphorylated residues. The detection of R1810me2s improves for both antibodies when the precipitated samples are treated with alkaline phosphatase. **d**, Slot blots illustrating that the Y12 and R1810me2s antibodies specifically recognize peptides containing RNAP II R1810me2s. The indicated amounts of biotin-labelled 7mer CTD peptides bracketing R1810 with no modification, Arg-me2a, and Arg-me2s were blotted before incubating with the R1810me2s or Y12 antibodies. **e**, Western blot confirming efficient PRMT5 knockdown, and RT-qPCR assay confirming efficient CARM1 knockdown for experiment of Fig. 1c. **f**, **g**, Whole-cell lysate western blot controls for Fig. 1c,d, respectively. **h–j**, Whole cell lysate western blot controls for Fig. 3c–e, respectively.



### Extended Data Figure 2 | Recognition of R1810me2s by SMN.

**a–c**, *In vitro* fluorescence polarization (FP) peptide binding assays.

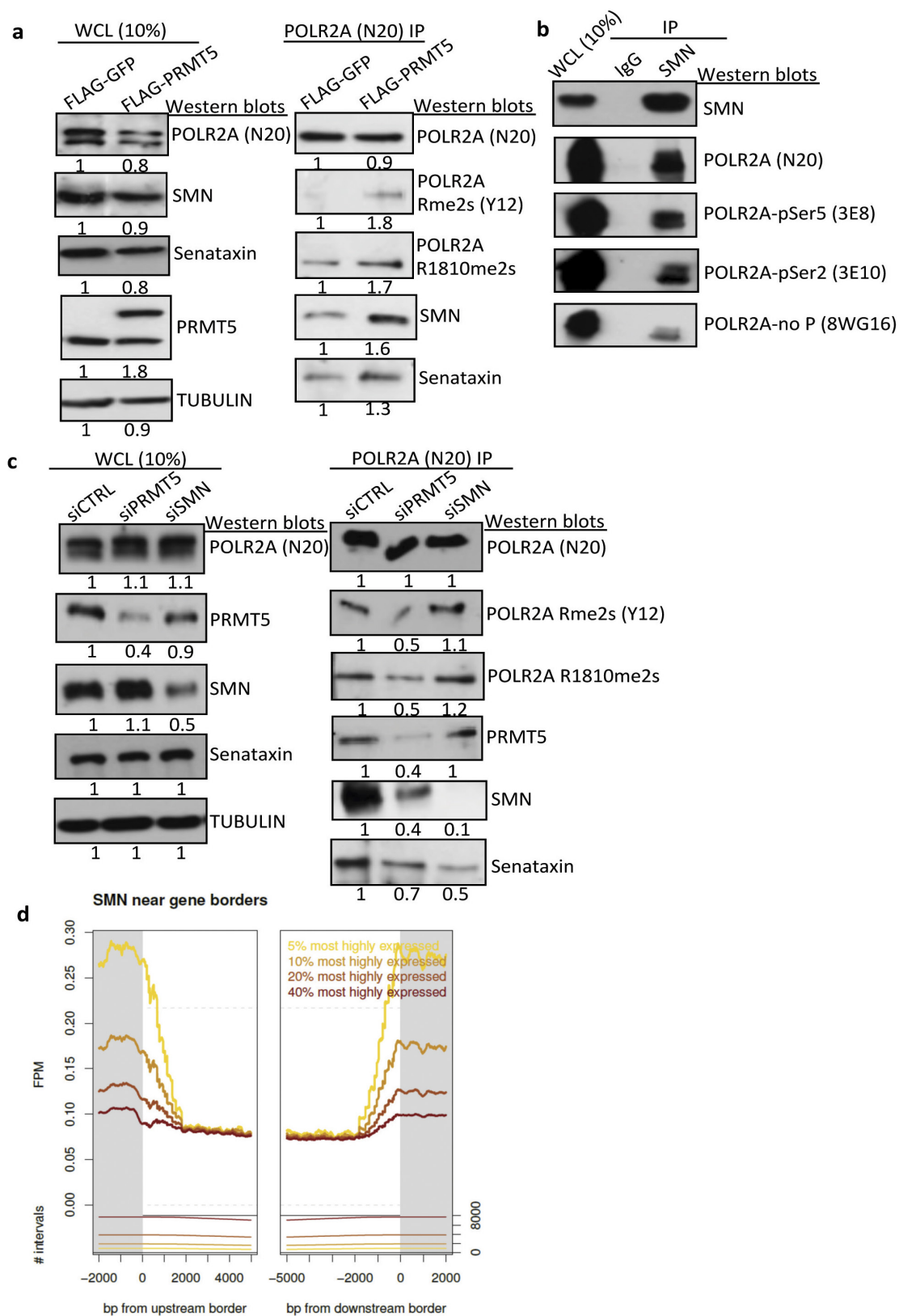
**a**, Recombinant SMN Tudor domain was incubated with FITC-labelled 13-mer CTD peptides bracketing either R1810 or R1603. SMN preferentially bound CTD peptides in the order R1810me2s > R1603me2s ~ R1810me2a > R1603me2a, and exhibited no detectable affinity for the unmodified peptides. **b**, Recombinant SMN Tudor domain was incubated with FITC-labelled CTD peptides bracketing R1810me2s also containing

Y1P, S2P or both upstream of R1810me2s, showing slightly preferential binding to the peptides when the phospho-modification(s) are present.

**c**, FITC-CTD R1810me2s or FITC-CTD R1810me2a is not recognized by other recombinant Tudor domains from SMNDC1 (also known as SPF30), TDRD1, TDRD2, TDRD9 or TDRD11 (also known as SND1).

**d**, Isothermal titration calorimetry assays showing that the recombinant SMN Tudor domain has no enhanced binding to R1810me2s containing peptides also carrying S2P or both Y1P and S2P.

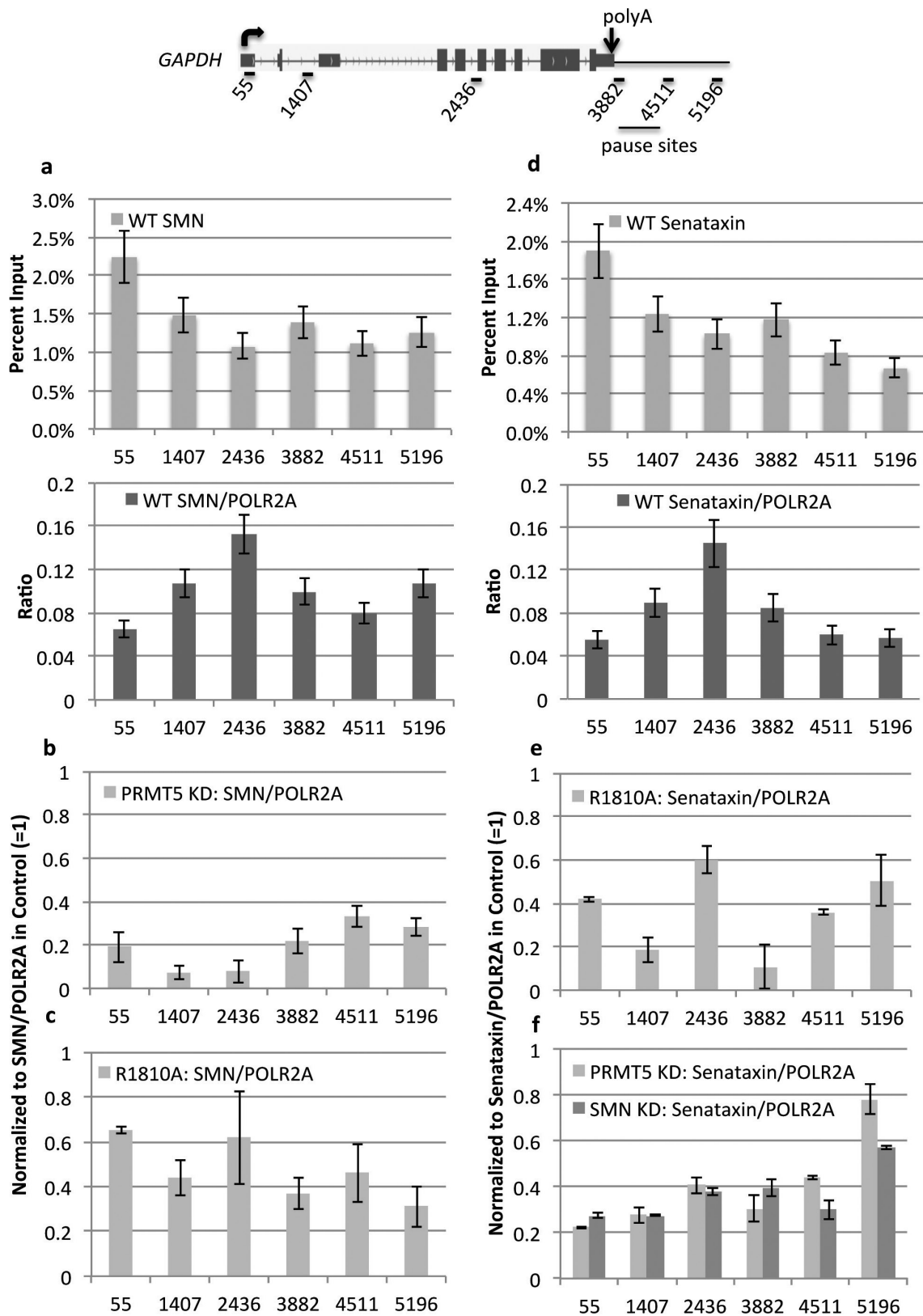




Extended Data Figure 3 | See next page for caption.

**Extended Data Figure 3 | Interaction of SMN and senataxin with RNAP II depends on PRMT5.** **a**, PRMT5 overexpression increases the R1810me2s modification and the SMN and senataxin associations with RNAP II. Left, western blots with the indicated antibodies of HEK293 whole-cell lysates overexpressing Flag-tagged PRMT5 or GFP. Overexpressing PRMT5 does not increase the amount of SMN or senataxin. Right, endogenous POLR2A was immunoprecipitated (N20 antibody) from HEK293 cell lysates with overexpressed Flag-tagged PRMT5 or GFP. **b**, SMN associates with phospho-isoforms of RNAP II. Immunoprecipitation with SMN antibody, but not control IgG, co-precipitated RNAP II with unmodified CTD repeats (8WG16 antibody) and CTD repeats phosphorylated on Ser5 and Ser2 as detected by western

blotting. **c**, Requirement of PRMT5 for association of SMN and senataxin with RNAP II. Left, western blots with the indicated antibodies of HEK293 whole-cell lysates expressing siRNAs for PRMT5 or SMN. Right, endogenous POLR2A was immunoprecipitated (N20 antibody) from cells with transient siRNA-mediated knockdown of PRMT5 or SMN. Western blots were performed with the indicated antibodies. PRMT5 knockdown causes loss of R1810me2s on POLR2A, as well as reduced association of SMN and senataxin with RNAP II. SMN knockdown causes reduced association of senataxin with RNAP II. **d**, SMN ChIP-seq (with GFP antibody against inducible GFP-SMN, or with SMN-specific antibody in HEK293 cells). Both methods observe enriched SMN signals at promoter and termination regions; the average of both is shown.



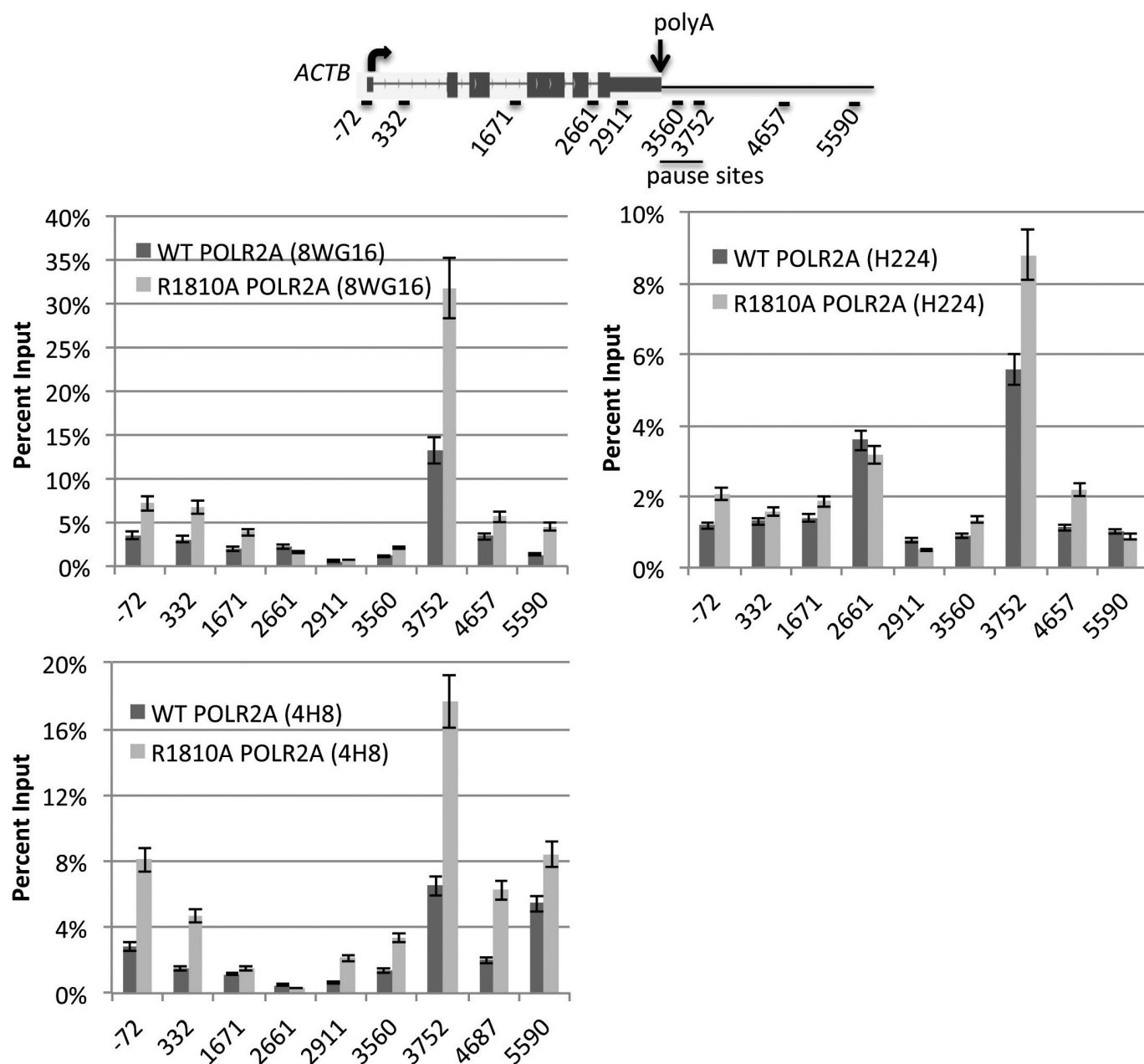
Extended Data Figure 4 | See next page for caption.



**Extended Data Figure 4 | Association of SMN and senataxin**

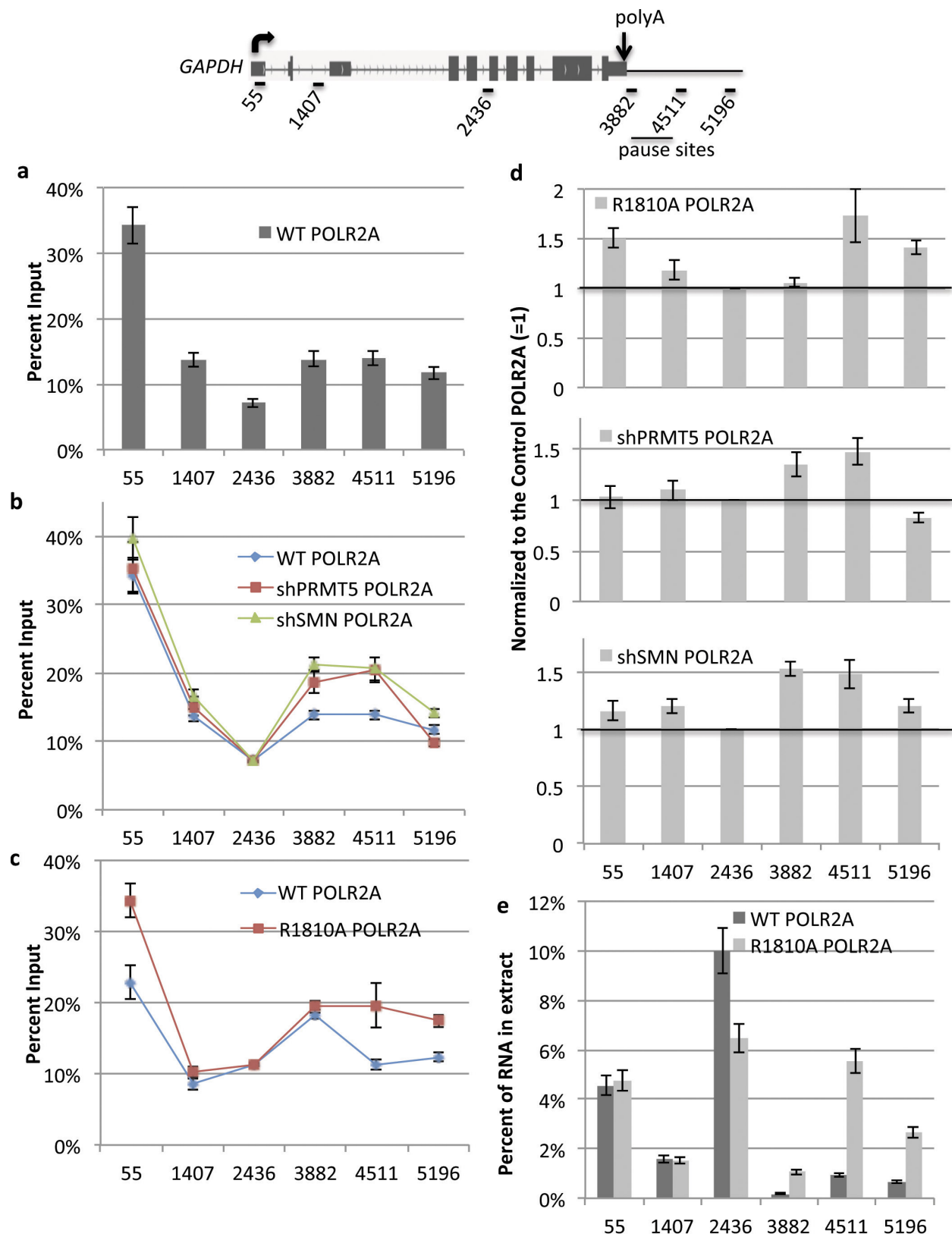
**with *GAPDH* depends on R1810 and PRMT5.** **a**, Chromatin immunoprecipitation (ChIP) was used to determine the distribution of SMN along the human *GAPDH* gene in HEK293 cells, expressed as percent input or as a ratio of SMN to RNAP II. Error bars represent technical variation in a single experiment (mean  $\pm$  s.e.m.,  $n = 3$ ). **b**, SMN ChIP was performed in HEK293 cells stably expressing shRNAs for PRMT5 or GFP (as a control). With the control normalized to 1, knockdown of PRMT5 caused strong reductions of the SMN ChIP signals all along *GAPDH*. Error bars represent biological replicates (mean  $\pm$  s.e.m.,  $n = 3$ ). **c**, SMN ChIP signals on *GAPDH* decrease in Raji cells expressing HA-tagged R1810A mutant POLR2A after 3 days of treatment with  $\alpha$ -amanitin to eliminate endogenous POLR2A. ChIP results with wild-type HA-tagged POLR2A were normalized to 1. Error bars represent biological replicates

(mean  $\pm$  s.e.m.,  $n = 3$ ). **d**, ChIP in HEK293 cells showing the distribution of senataxin along the human *GAPDH* gene, expressed as percent input or as a ratio to RNAP II. Error bars represent technical variation in a single experiment (mean  $\pm$  s.e.m.,  $n = 3$ ). **e**, Senataxin ChIP signals on *GAPDH* decrease in the POLR2A (R1810A) mutant after 3 days of treatment of Raji cells with  $\alpha$ -amanitin to eliminate endogenous POLR2A. Results are normalized to wild-type POLR2A and expressed as the ratio of senataxin to RNAP II. Error bars represent biological replicates (mean  $\pm$  s.e.m.,  $n = 3$ ). **f**, Knockdown of PRMT5 or SMN causes reductions of the senataxin ChIP signals all along *GAPDH*. ChIP against senataxin was performed in HEK293 cells with shRNA-mediated knock-down of PRMT5 or SMN. Results were normalized to a control knockdown of GFP and expressed as the ratio of senataxin to RNAP II. Error bars represent biological replicates (mean  $\pm$  s.e.m.,  $n = 2$ ).



**Extended Data Figure 5 | The R1810 mutation causes RNAP II to accumulate in the termination region of *ACTB*.** Chromatin immunoprecipitation (ChIP) with three different POLR2A antibodies (8WG16, H224, 4H8), as indicated, was performed on the *ACTB* gene in

Raji cells expressing HA-tagged wild-type or mutant (R1810A) POLR2A, 3 days after treatment with  $\alpha$ -amanitin to eliminate endogenous POLR2A. Shown are single experiments with error bars representing technical variation (mean  $\pm$  s.e.m.,  $n = 3$ ).



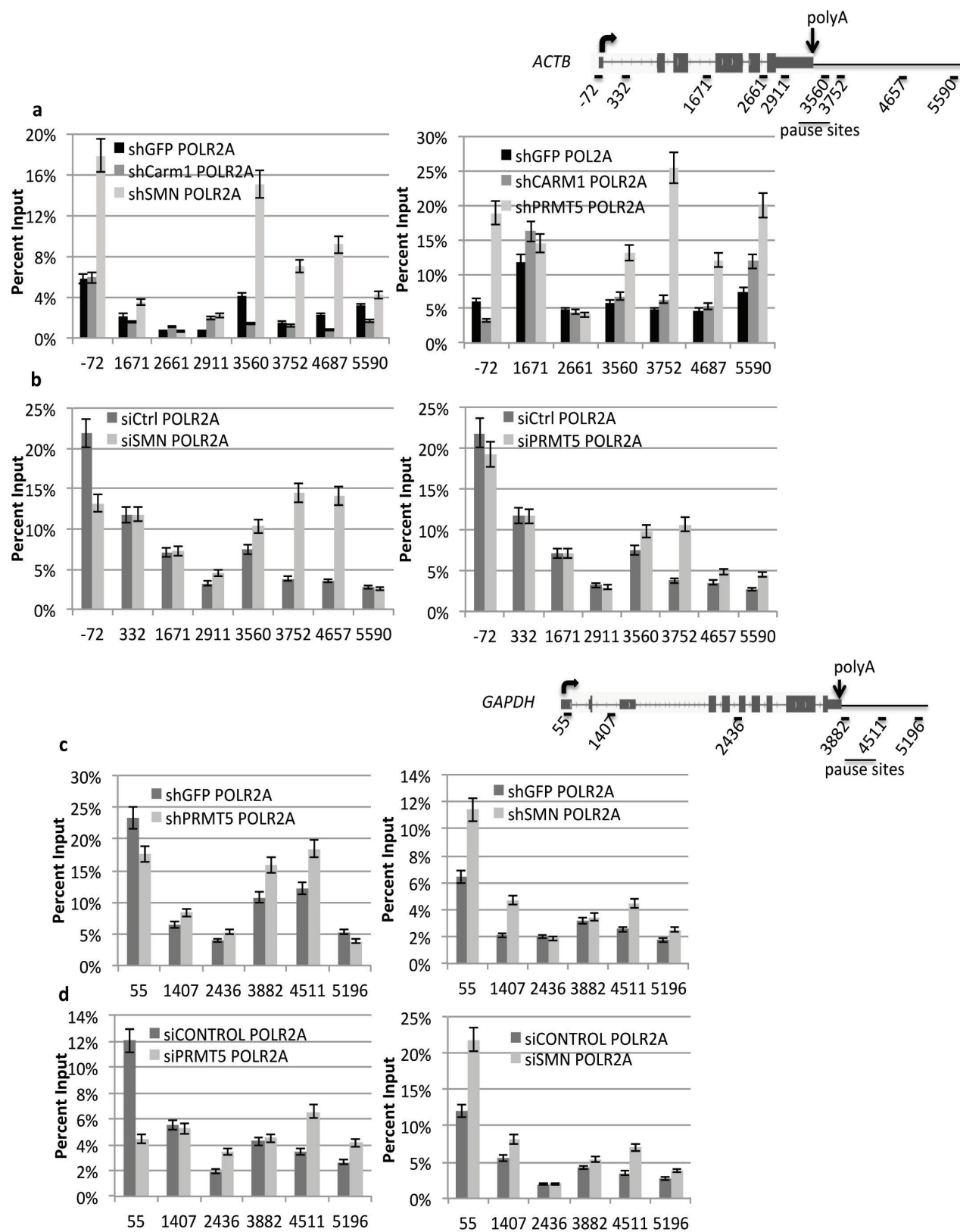
Extended Data Figure 6 | See next page for caption.



**Extended Data Figure 6 | R1810, PRMT5, and SMN regulate**

**transcription termination on *GAPDH*.** **a**, Chromatin immunoprecipitation (ChIP) with the N20, 4H8, and 8WG16 antibodies to show the distribution of wild-type POLR2A along the human *GAPDH* gene. Error bars represent technical variation (mean  $\pm$  s.e.m.,  $n = 3$ ). **b**, POLR2A ChIP along the *GAPDH* gene was performed in HEK293 cells after stable knockdown of PRMT5 or SMN, using stable knockdown of GFP as a negative control. RNAP II over-accumulates at the termination sites on *GAPDH* after knockdown of PRMT5 or SMN. Error bars represent biological replicates (mean  $\pm$  s.e.m.,  $n = 5$ ). **c**, POLR2A ChIP on the *GAPDH* gene was performed in Raji cells that express HA-tagged wild-type or mutant (R1810A) POLR2A 3 days after  $\alpha$ -amanitin treatment to eliminate endogenous POLR2A. R1810A mutant RNAP II over-accumulates

downstream of the cleavage and polyadenylation sites where RNAP II pauses and terminates transcription on *GAPDH*. Error bars represent biological replicates (mean  $\pm$  s.e.m.,  $n = 4$ ). **d**, Data from **b**, **c** are displayed as normalized ratios to the control (GFP knockdown, or HA wild-type POLR2A), with the ratio for the intron 5 qPCR primers at 2436 set as 1. **e**, Nuclear run-on experiment in which nuclei from Raji cells expressing wild-type or mutant (R1810A) POLR2A 3 days after  $\alpha$ -amanitin treatment to eliminate endogenous POLR2A were incubated with BrUTP for 30 min, and short run-on RNAs were isolated by binding to anti-BrU antibodies. The R1810A mutation led to over-accumulation of active RNAP II in the region downstream of the poly(A) site on *GAPDH*. Error bars represent technical variation (mean  $\pm$  s.e.m.,  $n = 3$ ).



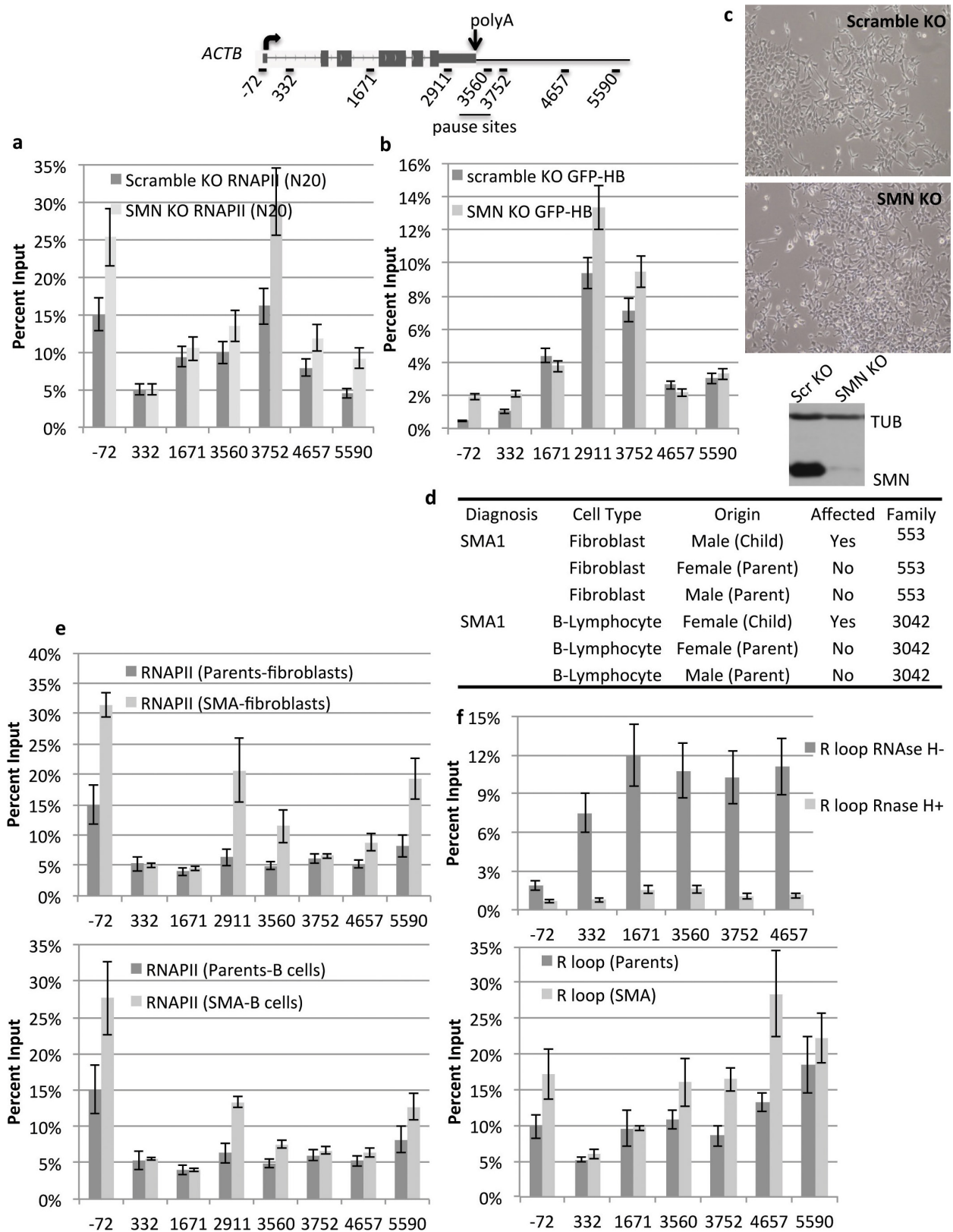
Extended Data Figure 7 | See next page for caption.

**Extended Data Figure 7 | PRMT5 and SMN but not CARM1 regulate transcription termination by RNAP II on *ACTB* and *GAPDH*.**

**a**, Chromatin immunoprecipitation (ChIP) for POLR2A with 4H8 antibody was performed after stable shRNA-mediated PRMT5, CARM1 or SMN knockdown in HEK293 cells to show that only PRMT5 and SMN knockdowns lead to the over-accumulation of RNAP II in the termination regions of *ACTB*. The graph shows a single experiment with error bars representing technical variation (mean  $\pm$  s.e.m.,  $n = 3$ ). **b**, POLR2A ChIP with the 8WG16 antibody was performed after transient siRNA-mediated knockdown of PRMT5 or SMN in HEK293 cells to show that PRMT5 or SMN knockdown leads to the over-accumulation of RNAP II in the termination region of  $\beta$ -actin. The graph shows a single experiment with

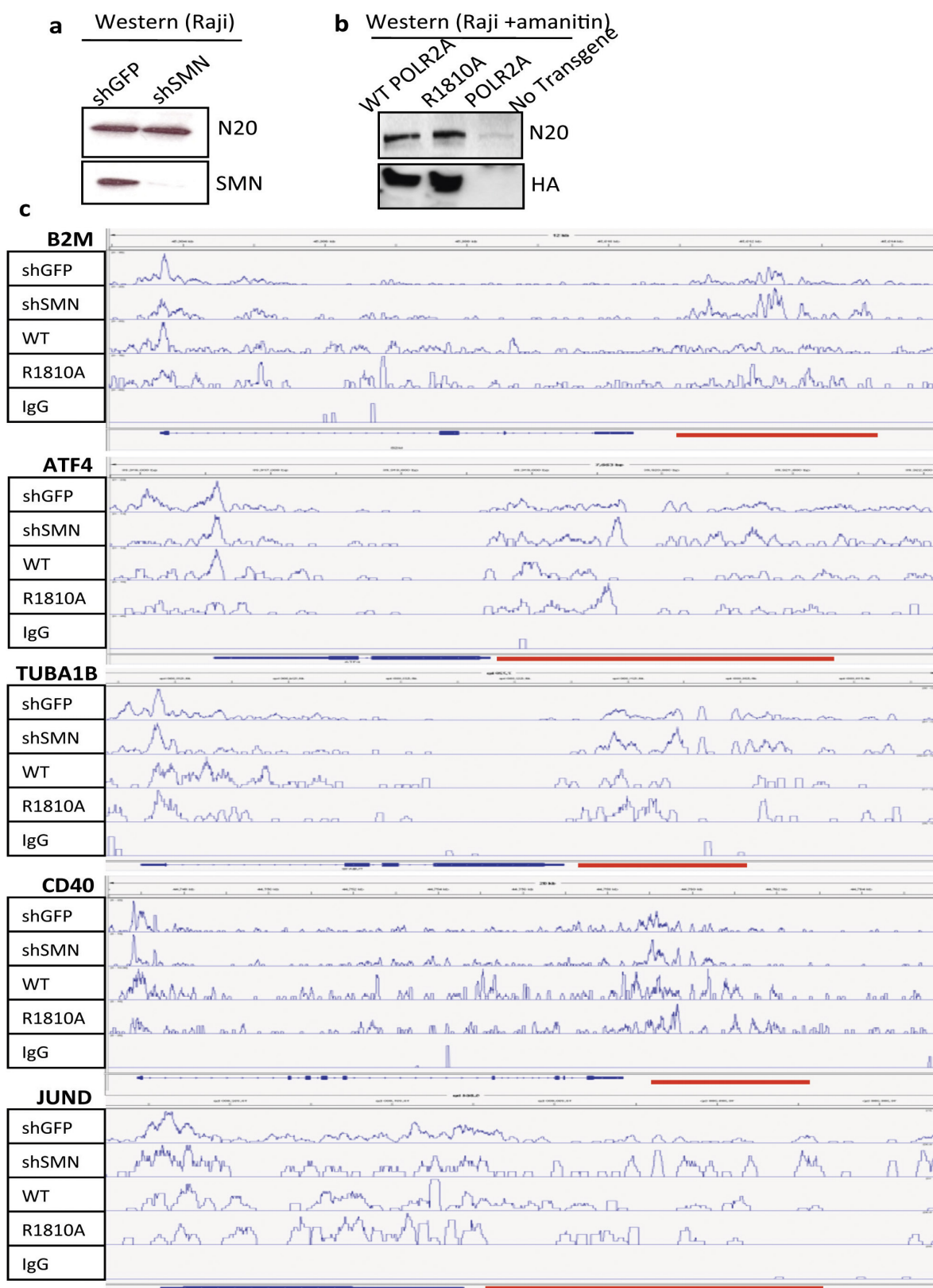
error bars representing technical variation (mean  $\pm$  s.e.m.,  $n = 3$ ). **c**, ChIP for POLR2A with 4H8 antibody was performed after stable shRNA-mediated knockdown of PRMT5, SMN or GFP (as a control) in HEK293 cells to show that knockdown of PRMT5 or SMN leads to the over-accumulation of RNAP II in the termination region of *GAPDH*. The graph shows a single experiment with error bars representing technical variation (mean  $\pm$  s.e.m.,  $n = 3$ ). **d**, POLR2A ChIP with 8WG16 antibody was performed after transient siRNA-mediated knockdown of PRMT5 or SMN in HEK293 cells to show that knockdown of PRMT5 or SMN leads to the over-accumulation of RNAP II in the termination region of *GAPDH*. The graph shows a single experiment with error bars representing technical variation (mean  $\pm$  s.e.m.,  $n = 3$ ).





**Extended Data Figure 8 | RNAP II pausing defect and R-loop accumulation are observed in CRISPR SMN knockout cells and in SMA disease cells.** **a**, ChIP for POLR2A with N20 antibody was performed after stable SMN knockout (CRISPR) in HEK293 cells shows that RNAP II accumulates in the termination regions of *ACTB*. Scrambled guide RNA treatment was used as a negative control. Error bars represent biological variation (mean  $\pm$  s.e.m.,  $n = 3$ ). **b**, Accumulation of R-loops in the termination regions of the *ACTB* gene after SMN knockout. A fusion protein of GFP–RNase H DNA–RNA hybrid binding domain was stably expressed in HEK293 cells. ChIP with GFP antibody (Abcam 290) was used for the detection of the R-loops (DNA–RNA hybrids), using the indicated primer positions for qPCR along the gene. Scrambled guide RNA treatment was used as a negative control. Error bars represent biological variation (mean  $\pm$  s.e.m.,  $n = 3$ ). **c**, Top: live cell microscopy images showing that HEK293 cells with SMN knockout appear to be physiologically normal in comparison to the control scrambled KO.

Bottom: western blot with anti-SMN antibody showing that *SMN* expression is knocked out by CRISPR. **d**, Human cell lines (3 fibroblast, 3 B lymphocyte) were obtained from the Coriell Institute for Medical Research. These include cells from two children with SMA disease and their normal parents. **e**, ChIP for POLR2A (N20, 8WG16 antibodies) was performed on the *ACTB* gene using the averaged value of the parents as control. POLR2A in the SMA disease cells (from both fibroblast and B cell lines) accumulates in the termination regions of the *ACTB* gene. Error bars represent biological variation (mean  $\pm$  s.e.m.,  $n = 4$ ). **f**, Top: quantification of R-loops by DNA immunoprecipitation (DIP) with the S9.6 antibody in the patient cells, showing that the R-loops are sensitive to RNase H. Error bars represent technical variation (mean  $\pm$  s.e.m.,  $n = 3$ ). Bottom: R-loop DIP with the S9.6 antibody shows that R-loops accumulate in the termination regions of the *ACTB* gene in the SMA disease cells. The averaged value of the parents was used as a control. Error bars represent biological variation (mean  $\pm$  s.e.m.,  $n = 5$ ).

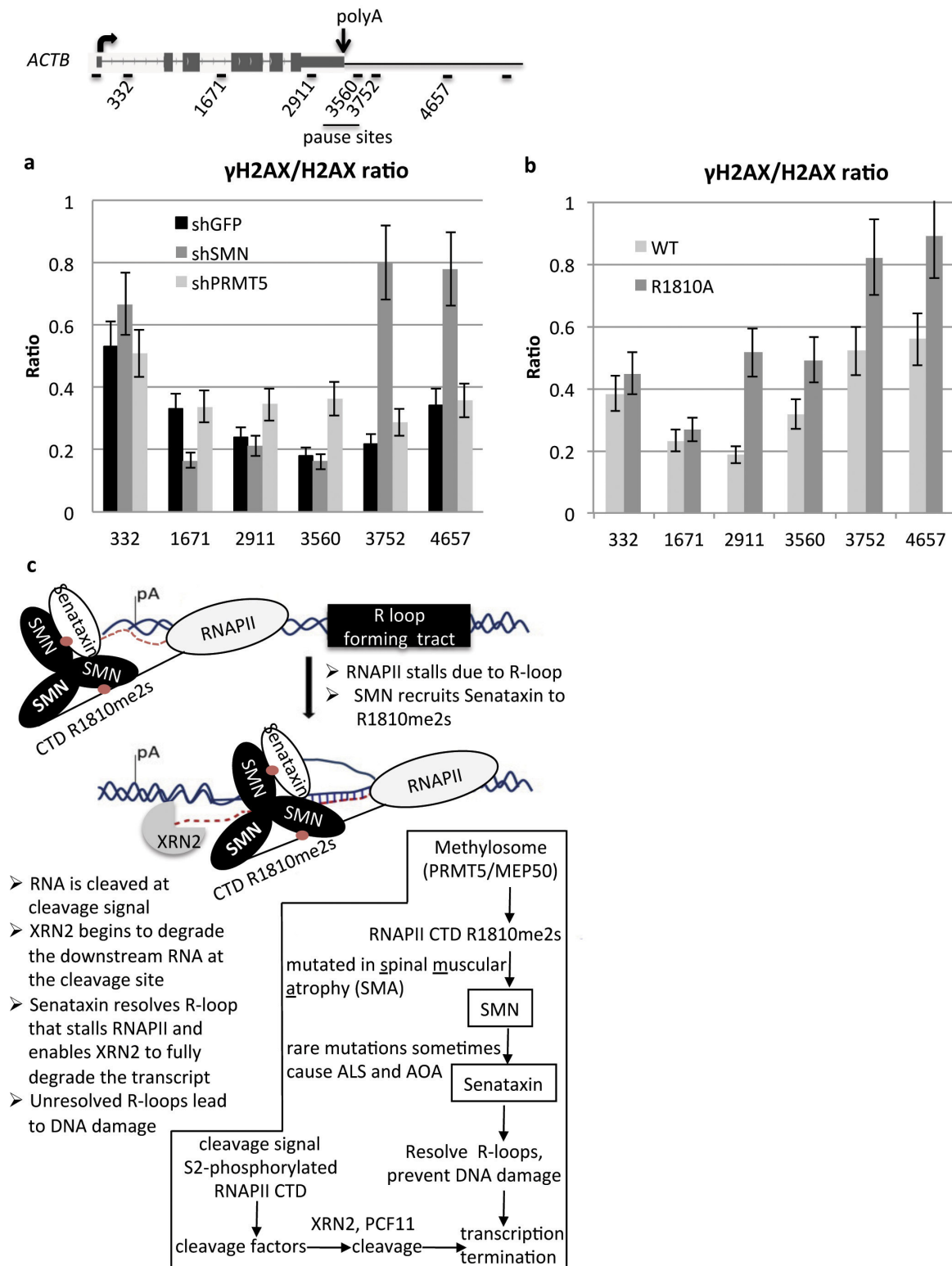


### Extended Data Figure 9 | SMN and POLR2A(R1810) effects on transcription termination by RNAP II occur on a genome-wide level.

**a**, Western blot using N20 antibody for POLR2A and SMN antibody to verify shSMN knockdown of SMN for the ChIP-seq experiment of Fig. 4e; shGFP was used as a control. **b**, Western blot using N20 and HA antibodies to verify equal expression of HA-tagged wild-type and POLR2A(R1810A) and the effect of  $\alpha$ -amanitin treatment on cells without an HA-tagged construct for the ChIP-seq experiment of Fig. 4e. **c**, RNAP II ChIP-seq results for several housekeeping genes are displayed in detail (B2M, CD40,

ATF4 (a short gene), JUND (an intronless gene), and TUBA1B) using the Integrative Genomics Viewer. The promoter peaks are displayed to the left, and the regions underlined in red are RNAP II termination regions. IgG ChIP-seq was used as negative control. Approximately 10 million unique RNAP II ChIP-seq reads (4H8, 8WG16) were obtained from GFP or SMN stable knockdown Raji cells. Approximately 10–12 million unique RNAP II ChIP-seq reads (N20) were obtained from wild-type or R1810A POLR2A Raji cells upon 3 days of amanitin treatment ( $2 \mu\text{g ml}^{-1}$ ).





**Extended Data Figure 10 | Model of the pathway that regulates R-loop accumulation to prevent DNA damage.** **a, b**, Quantification of  $\gamma$ H2AX:H2AX ratio through ChIP in HEK293 (**a**) or Raji (**b**) cells, along the *ACTB* gene, after knocking down PRMT5, or SMN, with GFP

knockdown as a negative control (**a**), or mutating R1810 to alanine (**b**). Error bars denote s.e.m. ( $n = 4$ ). **c**, Pathway (boxed) and model showing influence of PRMT5, R1810me2s, and SMN on R-loop resolution and transcription termination.

# Repetitive patterns in rapid optical variations in the nearby black-hole binary V404 Cygni

Mariko Kimura<sup>1</sup>, Keisuke Isogai<sup>1</sup>, Taichi Kato<sup>1</sup>, Yoshihiro Ueda<sup>1</sup>, Satoshi Nakahira<sup>2</sup>, Megumi Shidatsu<sup>3</sup>, Teruaki Enoto<sup>1,4</sup>, Takafumi Hori<sup>1</sup>, Daisaku Nogami<sup>1</sup>, Colin Littlefield<sup>5</sup>, Ryoko Ishioka<sup>6</sup>, Ying-Tung Chen<sup>6</sup>, Sun-Kun King<sup>6</sup>, Chih-Yi Wen<sup>6</sup>, Shiang-Yu Wang<sup>6</sup>, Matthew J. Lehner<sup>6,7,8</sup>, Megan E. Schwamb<sup>6</sup>, Jen-Hung Wang<sup>6</sup>, Zhi-Wei Zhang<sup>6</sup>, Charles Alcock<sup>8</sup>, Tim Axelrod<sup>9</sup>, Federica B. Bianco<sup>10</sup>, Yong-Ik Byun<sup>11</sup>, Wen-Ping Chen<sup>12</sup>, Kem H. Cook<sup>6</sup>, Dae-Won Kim<sup>13</sup>, Typhoon Lee<sup>6</sup>, Stuart L. Marshall<sup>14</sup>, Elena P. Pavlenko<sup>15</sup>, Oksana I. Antonyuk<sup>15</sup>, Kirill A. Antonyuk<sup>15</sup>, Nikolai V. Pit<sup>15</sup>, Aleksei A. Sosnovskij<sup>15</sup>, Julia V. Babina<sup>15</sup>, Aleksei V. Baklanov<sup>15</sup>, Alexei S. Pozanenko<sup>16,17</sup>, Elena D. Mazaeva<sup>16</sup>, Sergei E. Schmalz<sup>18</sup>, Inna V. Reva<sup>19</sup>, Sergei P. Belan<sup>15</sup>, Raguli Ya. Inasaridze<sup>20</sup>, Namkhair Tungalag<sup>21</sup>, Alina A. Volnova<sup>16</sup>, Igor E. Molotov<sup>22</sup>, Enrique de Miguel<sup>23,24</sup>, Kiyoshi Kasai<sup>25</sup>, William L. Stein<sup>26</sup>, Pavol A. Dubovsky<sup>27</sup>, Seiichiro Kiyota<sup>28</sup>, Ian Miller<sup>29</sup>, Michael Richmond<sup>30</sup>, William Goff<sup>31</sup>, Maksim V. Andreev<sup>32,33</sup>, Hiromitsu Takahashi<sup>34</sup>, Naoto Kojiguchi<sup>35</sup>, Yuki Sugiura<sup>35</sup>, Nao Takeda<sup>35</sup>, Eiji Yamada<sup>35</sup>, Katsura Matsumoto<sup>35</sup>, Nick James<sup>36</sup>, Roger D. Pickard<sup>37,38</sup>, Tamás Tordai<sup>39</sup>, Yutaka Maeda<sup>40</sup>, Javier Ruiz<sup>41,42,43</sup>, Atsushi Miyashita<sup>44</sup>, Lewis M. Cook<sup>45</sup>, Akira Imada<sup>46</sup> & Makoto Uemura<sup>47</sup>

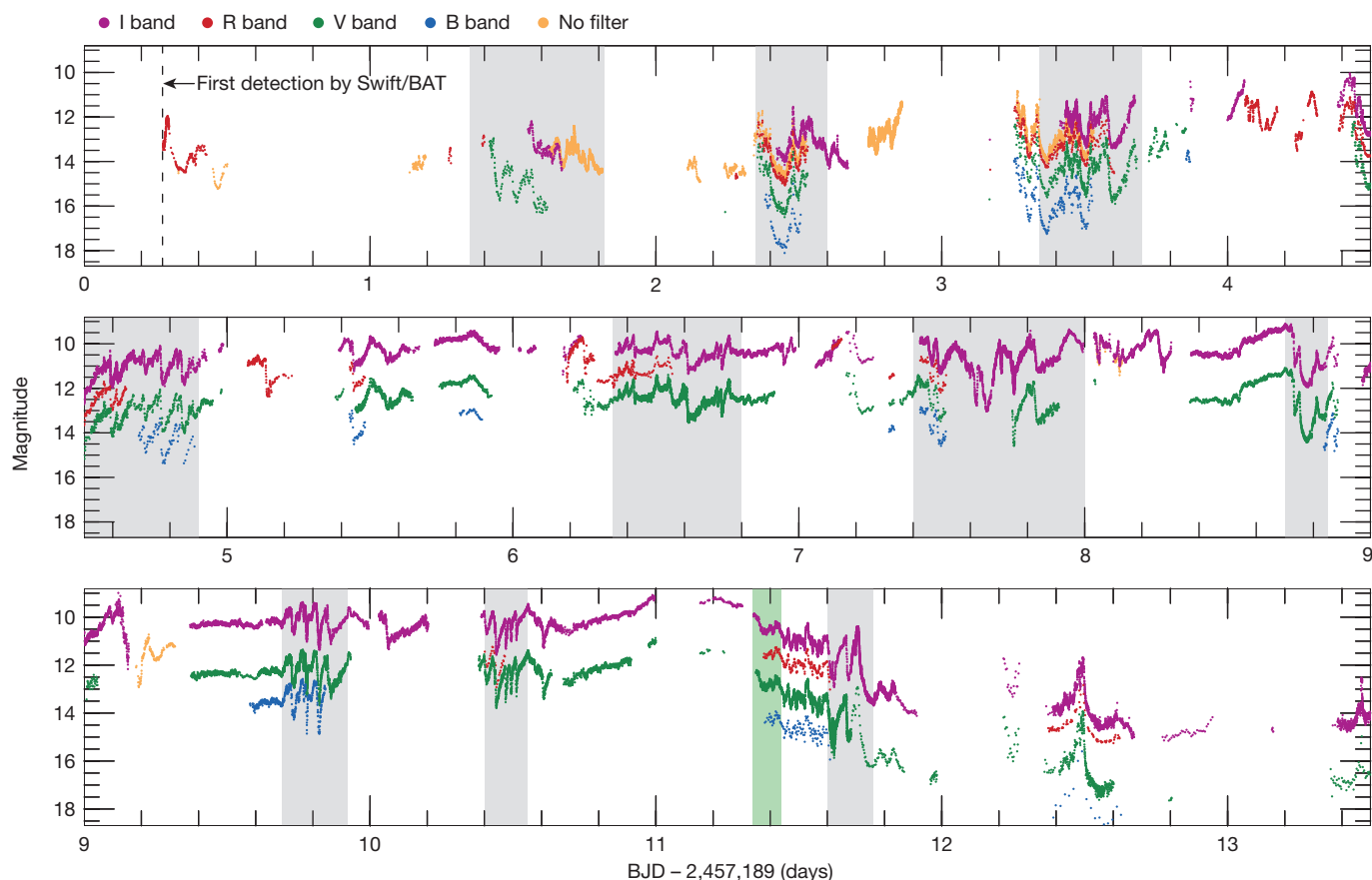
How black holes accrete surrounding matter is a fundamental yet unsolved question in astrophysics. It is generally believed that matter is absorbed into black holes via accretion disks, the state of which depends primarily on the mass-accretion rate. When this rate approaches the critical rate (the Eddington limit), thermal instability is supposed to occur in the inner disk, causing repetitive patterns of large-amplitude X-ray variability (oscillations) on timescales of minutes to hours<sup>1</sup>. In fact, such oscillations have been observed only in sources with a high mass-accretion rate, such as GRS 1915+105 (refs 2, 3). These large-amplitude, relatively slow timescale, phenomena are thought to have physical origins distinct from those of X-ray or optical variations with small amplitudes and fast timescales (less than about 10 seconds) often observed in other black-hole binaries—for example, XTE J1118+480 (ref. 4) and GX 339–4 (ref. 5). Here we report an extensive multi-colour optical photometric data set of V404 Cygni, an X-ray transient source<sup>6</sup> containing a black hole of nine solar masses<sup>7</sup> (and a companion star) at a distance of 2.4 kiloparsecs (ref. 8). Our data show that optical oscillations on timescales of 100 seconds to 2.5 hours can occur at mass-accretion rates more than ten times lower than previously thought<sup>1</sup>. This suggests that the accretion rate is not the critical parameter for inducing inner-disk instabilities. Instead, we propose

that a long orbital period is a key condition for these large-amplitude oscillations, because the outer part of the large disk in binaries with long orbital periods will have surface densities too low to maintain sustained mass accretion to the inner part of the disk. The lack of sustained accretion—not the actual rate—would then be the critical factor causing large-amplitude oscillations in long-period systems.

V404 Cyg, which was originally discovered as a nova in 1938 and detected by the GINGA satellite in 1989<sup>9</sup>, underwent an outburst in June 2015 after 26 years of dormancy. At 18:31:38 on June 15 (15.77197 Universal Time (UT)), Swift/Burst Alert Telescope (BAT) initially detected this outburst as a possible  $\gamma$ -ray burst<sup>10</sup>. The outburst was also detected by the Monitor of All-sky X-ray Image (MAXI) instrument on June 16.783 UT<sup>11</sup>.

Following these detections, we started a world-wide photometric campaign (Extended Data Tables 1 and 2, Methods section ‘Detailed methods of optical observations’) partly within the Variable Star Network (VSNET) Collaboration, and collected extensive sets of multi-colour optical photometric data consisting of >85,000 points. Our data set also includes early observations with the Taiwanese–American Occultation Survey (TAOS) starting on June 15, 18:34:07 UT, 2 min 29 s after the Swift/BAT trigger<sup>12</sup> (see Extended Data Tables 1 and 2, and Methods section ‘Detailed methods of optical observations’ concerning

<sup>1</sup>Department of Astronomy, Graduate School of Science, Kyoto University, Oiwakecho, Kitashirakawa, Sakyo-ku, Kyoto 606-8502, Japan. <sup>2</sup>JEM Mission Operations and Integration Center, Human Spaceflight Technology Directorate, Japan Aerospace Exploration Agency, 2-1-1 Sengen, Tsukuba, Ibaraki 305-8505, Japan. <sup>3</sup>MAXI team, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan. <sup>4</sup>The Hakubi Center for Advanced Research, Kyoto University, Kyoto 606-8302, Japan. <sup>5</sup>Astronomy Department, Wesleyan University, Middletown, Connecticut 06459, USA. <sup>6</sup>Institute of Astronomy and Astrophysics, Academia Sinica, 11F of Astronomy-Mathematics Building, AS/NTU No. 1, Section 4, Roosevelt Road, Taipei 10617, Taiwan. <sup>7</sup>Department of Physics and Astronomy, University of Pennsylvania, 209 South 33rd Street, Philadelphia, Pennsylvania 19125, USA. <sup>8</sup>Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. <sup>9</sup>Steward Observatory, University of Arizona, Tucson, Arizona 85721, USA. <sup>10</sup>Center for Cosmology and Particle Physics, New York University, 4 Washington Place, New York, New York 10003, USA. <sup>11</sup>Department of Astronomy and University Observatory, Yonsei University, Seoul 120-749, South Korea. <sup>12</sup>Institute of Astronomy and Department of Physics, National Central University, Chung-Li 32054, Taiwan. <sup>13</sup>Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany. <sup>14</sup>Kavli Institute for Particle Astrophysics and Cosmology (KIPAC), Stanford University, 452 Lomita Mall, Stanford, California 94309, USA. <sup>15</sup>Crimean Astrophysical Observatory, 298409 Nauchny, Crimea. <sup>16</sup>Space Research Institute, Russian Academy of Sciences, 117997 Moscow, Russia. <sup>17</sup>National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Moscow, Russia. <sup>18</sup>Leibniz Institute for Astrophysics, Potsdam, Germany. <sup>19</sup>Fesenkov Astrophysical Institute, Almaty, Kazakhstan. <sup>20</sup>Kharadze Abastumani Astrophysical Observatory, Iliia State University, Tbilisi, Georgia. <sup>21</sup>Institute of Astronomy and Geophysics, Mongolian Academy of Sciences, Ulaanbaatar 13343, Mongolia. <sup>22</sup>Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Moscow, Russia. <sup>23</sup>Departamento de Física Aplicada, Facultad de Ciencias Experimentales, Universidad de Huelva, 21071 Huelva, Spain. <sup>24</sup>Center for Backyard Astrophysics, Observatorio del CIECEM, Parque Dunar, Matalascañas, 21760 Almonte, Huelva, Spain. <sup>25</sup>Baselstrasse 133D, CH-4132 Muttetzn, Switzerland. <sup>26</sup>6025 Calle Paraiso, Las Cruces, New Mexico 88012, USA. <sup>27</sup>Vihorlat Observatory, Mierova 4, Humenne, Slovakia. <sup>28</sup>Variable Star Observers League in Japan (VSOLJ), 7-1 Kitahatsutomi, Kamagaya, Chiba 273-0126, Japan. <sup>29</sup>Furzehill House, Ilston, Swansea SA2 7LE, UK. <sup>30</sup>Physics Department, Rochester Institute of Technology, Rochester, New York 14623, USA. <sup>31</sup>American Association of Variable Star Observers (AAVSO), 13508 Monitor Lane, Sutter Creek, California 95685, USA. <sup>32</sup>Institute of Astronomy, Russian Academy of Sciences, 361605 Peak Terskol, Kabardino-Balkaria, Russia. <sup>33</sup>International Center for Astronomical, Medical and Ecological Research of National Academy of Sciences of Ukraine (NASU), 27 Akademika Zabolotnoho street, 03680 Kiev, Ukraine. <sup>34</sup>Department of Physical Science, School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan. <sup>35</sup>Osaka Kyoiku University, 4-698-1 Asahigaoka, Kashiwara, Osaka 582-8582, Japan. <sup>36</sup>1 Tavistock Road, Chelmsford, Essex CM1 6JL, UK. <sup>37</sup>The British Astronomical Association, Variable Star Section (BAA VSS), Burlington House, Piccadilly, London W1J 0DU, UK. <sup>38</sup>The Birches, Shobdon, Leominster, Herefordshire HR6 9NG, UK. <sup>39</sup>Polaris Observatory, Hungarian Astronomical Association, Laborc utca 2/c, 1037 Budapest, Hungary. <sup>40</sup>112-14 Kaminishiyama-machi, Nagasaki, Nagasaki 850-0006, Japan. <sup>41</sup>Observatorio de Cantabria, Carretera de Rocamundo sin número, Valderredible, Cantabria, Spain. <sup>42</sup>Instituto de Física de Cantabria (CSIC-UC), Avenida Los Castros sin número, E-39005 Santander, Cantabria, Spain. <sup>43</sup>Agrupación Astronómica Cantabra, Apartado 573, 39080 Santander, Spain. <sup>44</sup>Seikei Meteorological Observatory, Seikei High School, Kichijoji-kitamachi 3-10-13, Musashino, Tokyo 180-8633, Japan. <sup>45</sup>Center for Backyard Astrophysics (Concord), 1730 Helix Court, Concord, California 94518, USA. <sup>46</sup>Kwasan and Hida Observatories, Kyoto University, Kitakasan-Ohrime-cho, Yamashina-ku, Kyoto 607-8471, Japan. <sup>47</sup>Hiroshima Astrophysical Science Center, Hiroshima University, Kagamiyama 1-3-1, Higashi-Hiroshima, Hiroshima 739-8526, Japan.



**Figure 1 | Overall multi-colour light curves during the 2015 outburst of V404 Cyg.** Shown are multi-colour light curves (B, V, R and I bands, and no filters) during BJD 2,457,189 to 2,457,202 (BJD 2,457,189 corresponds to 2015 June 15). It is clearly seen that dip-type oscillations (variations with recurrent sudden dips) were observed from the beginning to the

end of the outburst. The horizontal axis shows BJD – 2,457,189. The significant periods of repetitive optical variations are indicated in grey and green shading for the ‘dip-type’ and ‘heartbeat-type’ oscillations, respectively.

the VSNET collaboration team and TAOS). Some weak activity started approximately 1,000 s before the Swift/BAT trigger<sup>13</sup>. The same activity above 80 keV was also detected by the active anti-coincidence shield (ACS) of the Spectrometer on INTEGRAL (SPI) telescope of the INTEGRAL observatory in the same time intervals (P. Minaev, personal communication).

Our observations immediately indicated that large-amplitude short-term variations on timescales of  $\sim 100$  s to  $\sim 2.5$  h were already present, starting less than three minutes after the Swift/BAT trigger. In Fig. 1 and Extended Data Fig. 1, we show the overall optical multi-colour light curves. The overall trend of the light curves can be divided into three stages: (1) gradual rise during BJD (Barycentric Julian Day) 2,457,189 to 2,457,194.5 (brightening by  $1 \text{ mag d}^{-1}$  on average); (2) the plateau during BJD 2,457,194.5 to 2,457,200.0; and (3) rapid fading during BJD 2,457,200.0 to 2,457,203.3 (fading on average by  $2.5 \text{ mag d}^{-1}$ ). Short-term variations with amplitudes varying between 0.1 mag and 2.5 mag were observed throughout the outburst, and consisted of characteristic structures such as recurrent sudden dips from a peak (Fig. 1).

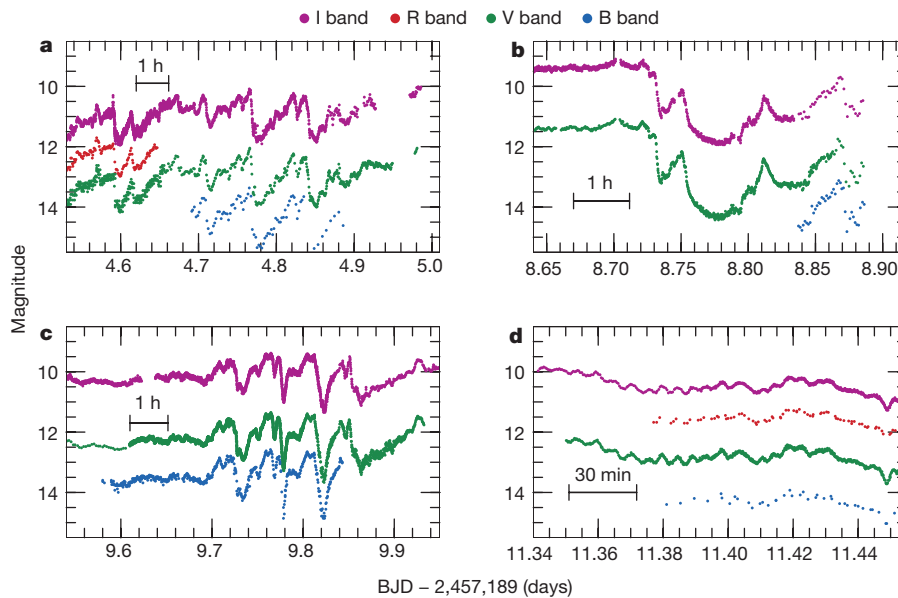
Moreover, fluctuations similar in shape to the unique X-ray variations of the enigmatic black-hole binary GRS 1915 + 105<sup>2</sup> are present in the optical light curve of V404 Cyg (Fig. 2). The patterns in the X-ray light curve of GRS 1915 + 105 have been classified into at least 12 categories on the basis of their flux and colour characteristics<sup>3</sup>. Repeating structures like these had not been observed in optical wavelengths before the 2015 outburst of V404 Cyg. The variations that we observed can be divided into two characteristic classes: (1) ‘dip-type’ oscillations (repetitions of a gradual rise followed by a sudden dip, sometimes with accompanying spikes on timescales of  $\sim 45$  min to  $\sim 2.5$  h; Fig. 2a–c); and (2) ‘heartbeat-type’ oscillations (rhythmic small spikes with short

periods of  $\sim 5$  min; Fig. 2d). Although rapid optical variations have been detected in the black-hole binary V4641 Sgr, those variations are stochastic with no indication of regular patterns<sup>14</sup>. The variations we found in V404 Cyg at optical wavelengths were regular and similar in shape to those in GRS 1915 + 105, although the interval between dips is about 5 times larger in V404 Cyg than in GRS 1915 + 105.

Using X-ray data from Swift/X-ray Telescope (XRT), we compared simultaneous optical and X-ray light curves (Fig. 3). When both X-ray and optical data showed strong short-term variations, the temporal correlations were generally good, although the X-ray flux variations are much larger than the optical ones. The good correlation indicates that both X-ray and optical observations recorded the same phenomena (see also Methods section ‘Comparison with X-ray observations’ and Extended Data Fig. 2). Spectral analyses of the simultaneous X-ray data (Methods section ‘Origin of cyclic dips’ and Extended Data Fig. 3) indicate that there was no tendency for increased absorption when the X-ray flux decreased, suggesting that these dips do not originate in absorption. In some epochs, we found evidence for heavy obscuration as found in the GINGA data during the 1989 outburst<sup>15</sup>; however this is not related to dip-type variations. We can thus infer that the short-term fluctuations directly reflect variations in radiation from the accretion disk or its associated structures. Detailed analyses of the typical simultaneous broad-band spectral energy distribution (SED) (Methods section ‘SED modelling’ and Extended Data Fig. 6) show that the majority of the optical flux is most likely to be produced by reprocessing of X-ray irradiation in the disk.

For GRS 1915 + 105, it has been proposed that the observed variability is caused by limit-cycle oscillations in the inner accretion disk due to Lightman–Eardley viscous instability<sup>16</sup>, which can explain a slow rise in



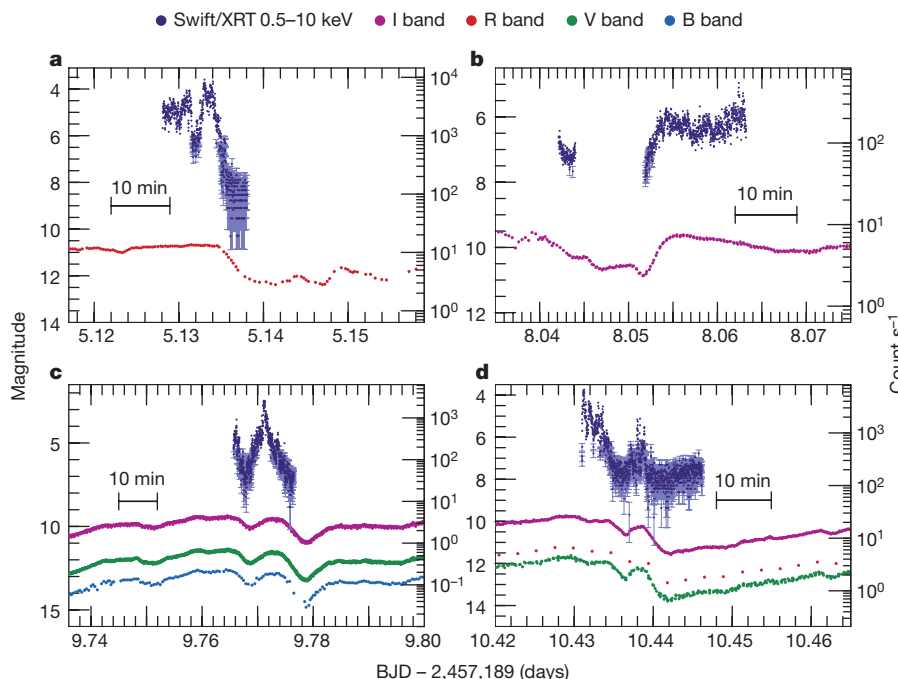


**Figure 2 | Short-term and large-amplitude optical variations having repeating structures in the 2015 outburst of V404 Cyg.** **a–d**, Variations with characteristic patterns during BJD 2,457,193.6 to 2,457,194.0 (**a**), BJD 2,457,197.7 to 2,457,198.0 (**b**), BJD 2,457,198.6 to 2,457,198.9 (**c**) and BJD 2,457,200.34 to 2,457,200.6 (**d**). In **a**, **b** and **c**, there are gradual rises with increasing amplitudes of fluctuations followed by dips,

brightness (mass accumulation) followed by a sudden drop (accretion to the black hole). Such a model assumes that the black hole is accreting mass nearly at the Eddington rate, which is supported by observations of GRS 1915 + 105<sup>17</sup>. Similar types of X-ray variability have also been detected in the black-hole binary IGR J17091–3624 (ref. 18), whose Eddington rate is unknown because both the mass and the distance are uncertain.

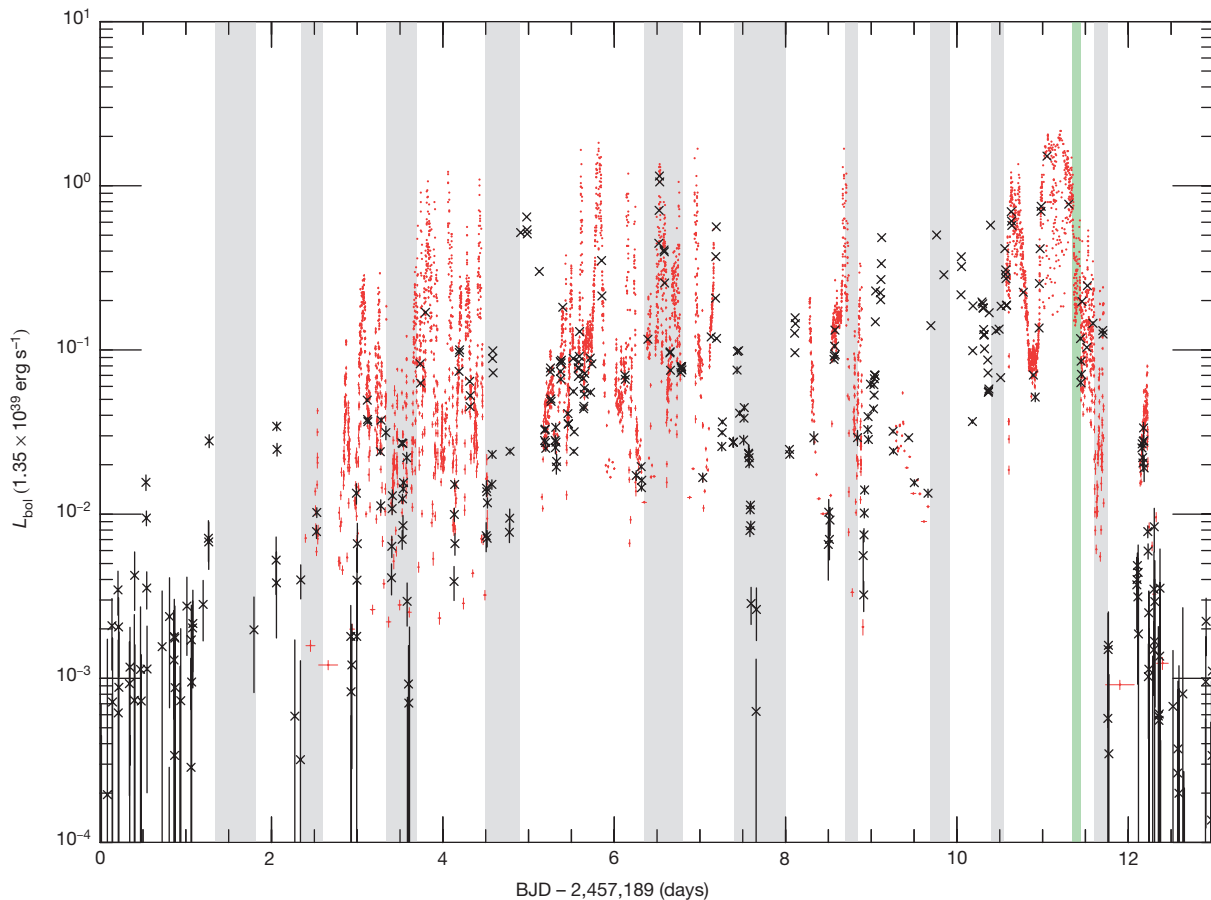
during which fluctuations disappear. These variations are sometimes accompanied by spikes. The interval between two dips ranges from ~45 min to ~2.5 h. **d**, Repetitive small oscillations with high coherence at intervals of ~5 min. The shapes of these oscillations resemble those of GRS 1915 + 105<sup>3</sup>.

In V404 Cyg, however, the accurate determination of the distance based on a parallax measurement<sup>8</sup> and the dynamical mass determination<sup>7</sup> enable us to conclude from our 2015 data that the black hole in this system was accreting at a much lower rate than the Eddington rate most of the time. During the period when GRS 1915 + 105-type variations in the optical light curves were recorded in V404 Cyg, its bolometric luminosity, averaged over an interval longer than the period of



**Figure 3 | Correlation between optical and X-ray fluctuations of V404 Cyg in the 2015 outburst.** The times covered in each panel are BJD 2,457,194.126 to 2,457,194.140 (**a**), BJD 2,457,197.050 to 2,457,197.065 (**b**), BJD 2,457,198.760 to 2,457,198.780 (**c**) and BJD 2,457,199.430 to 2,457,199.450 (**d**). In each panel, the left-hand y axis shows magnitude in bands I, R, V and B, and the right-hand y axis shows counts per second

in the Swift/XRT 0.5–10 keV band. Panels **a** and **b** cover the fading and rising phases, respectively; panels **c** and **d** show the correlations of short-term fluctuations. When both X-ray and optical light strongly varied, the correlation is generally good (though note in **a**, **c** and **d** that optical dips lag behind X-ray dips). Navy blue error bars,  $\pm 1\sigma$ . We plot points without errors when errors are smaller than or comparable to the plotting symbols.



**Figure 4 | The bolometric luminosity  $L_{\text{bol}}$  of V404 Cyg during the 2015 outburst.** It is normalized to the Eddington luminosity assuming a black hole mass of  $9M_{\odot}$ . Black points, Swift/BAT survey data (15–50 keV); red points, from the public Target Opportunity release of INTEGRAL Imager

on Board the Integral Satellite (IBIS)/CdTe array (ISGRI) monitoring (25–60 keV). Grey and green shadings represent respectively the periods of the ‘dip-type oscillations’ and the ‘heartbeat-type oscillations’. Black and red error bars,  $\pm 1\sigma$ .

oscillation, spanned a wide range, from  $\sim 0.01 L_{\text{Edd}}$  to  $\sim 0.4 L_{\text{Edd}}$  (where  $L_{\text{Edd}}$  is the Eddington luminosity for a nine solar-mass,  $M_{\odot}$ , black hole), as estimated from the hard X-ray flux and SED (Fig. 4 and Methods section ‘Time history of the bolometric luminosity’). Remarkably, the dip-type oscillations were observed at mean bolometric luminosity of  $\sim 0.015 L_{\text{Edd}}$ ,  $\sim 0.07 L_{\text{Edd}}$  and  $\sim 0.06 L_{\text{Edd}}$  during BJD 2,457,191.35 to 2,457,191.60, BJD 2,457,192.34 to 2,457,192.70, and BJD 2,457,200.60 to 2,457,200.76, respectively.

It is also worth noting that a typical dip similar to those seen in GRS 1915+105 was observed just 3 min after the first detection of this outburst (Extended Data Fig. 1b). This fact suggests that the accretion rate is not the critical parameter for inducing these oscillations. Our results imply that there is a novel type of disk instability that is different from the known dwarf-nova type<sup>19</sup> or the Lightman–Eardley type<sup>16</sup>.

We point out that black-hole binaries showing large-amplitude, short-term variations either in X-ray or optical bands have long orbital periods (33.9 d in GRS 1915+105<sup>20</sup>,  $\sim 4$  d in IGR J17091–3624<sup>21</sup>, 6.5 d in V404 Cyg<sup>22</sup>, and 2.8 d in V4641 Sgr<sup>23</sup>; see Methods section ‘Objects showing violent short-term variations in outburst’ and Extended Data Table 3 for a comparison of these objects), reinforcing the earlier suggestion of this link between violent oscillations and long orbital periods<sup>24</sup>. It has been proposed that the accretion disk in a system with a long orbital period suffers from instabilities in the disk’s vertical structure, and hence the disk beyond this radius of instability may never build up<sup>15,25</sup>. Our SED modelling of this outburst, however, requires a disk having a large radius ( $\gtrsim 1.7 \times 10^{12}$  cm), even considering the uncertainty of the interstellar reddening, to account for the ultraviolet flux particularly. This result implies that the disk extended up to distances close to the maximum achievable

radius (Methods section ‘SED modelling’). This radius is consistent with the short-term optical variations detected below 0.01 Hz (Extended Data Fig. 5 and Methods section ‘Power spectra’) and the time lag of  $\sim 1$  min between the X-ray and optical light curves (Fig. 3 and Extended Data Fig. 2) if we assume that the optical light mainly comes from reprocessed X-rays. We note that synchrotron emission has been proposed to be the origin of the short-term and large-amplitude fluctuations in the case of V4641 Sgr<sup>14</sup>. The optical polarization of V404 Cyg, however, did not show evidence of significant variations during the 2015 outburst<sup>26,27</sup>. This fact disfavours synchrotron emission as the origin of the short-term variations.

Outbursts of X-ray transients are thought to be triggered by the dwarf-nova-type instability: once the surface density at some radius reaches the critical density ( $\Sigma_{\text{crit}}$ ) after continuous mass transfer from the secondary star, thermal instability occurs and the disk undergoes an outburst<sup>19</sup>. In systems with long orbital periods, it is difficult for surface densities in the outer disk to reach  $\Sigma_{\text{crit}}$ , which is roughly proportional to the radius<sup>28</sup>. As a result, thermal instability in the inner part of the disk occurs more easily and governs the outburst behaviour<sup>29</sup>. This is probably the reason why long-period systems behave differently from short-period ‘classical’ X-ray transients. In fact, our estimate of the disk mass ( $5 \times 10^{25}$  g) accreted during the 2015 outburst is far smaller than the mass ( $2 \times 10^{28}$  g) of a fully built-up disk in quiescence (Methods section ‘Estimation of the disk mass and comparison with previous outbursts’ and Extended Data Fig. 4). These values indicate that the surface density was well below the  $\Sigma_{\text{crit}}$  required to induce thermal instability in most parts of the disk at the onset of the present outburst. Once the X-ray outburst started in the inner region, hydrogen atoms in the outer part of the disk would have been ionized and ‘passively’

maintained in the hot state as long as the X-ray illumination continued. This explains the large optical fluxes observed<sup>28</sup>. The rapid decay observed in the 2015 outburst of V404 Cyg may reflect the lack of the exponential decay in long-period systems as theoretically predicted<sup>30</sup>. Because the surface densities in the rest of the disk were too low to sustain the outburst by viscous diffusion<sup>19</sup>, only the inner part of the disk was responsible for the dynamics of the present outburst, as inferred from the rapid fading from the outburst (Methods section ‘Disk radius inferred from final fading rate’). We infer that, in outbursts of IGR J17091–3624<sup>18,21</sup> and the 1938 outburst of V404 Cyg (Methods section ‘Estimation of the disk mass and comparison with previous outbursts’ and Extended Data Fig. 4), the radius of the active disk is larger, which explains why the duration of those events is longer than that of the 2015 outburst of V404 Cyg.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 25 July; accepted 13 November 2015.**

- Janiuk, A. & Czerny, B. On different types of instabilities in black hole accretion discs: implications for X-ray binaries and active galactic nuclei. *Mon. Not. R. Astron. Soc.* **414**, 2186–2194 (2011).
- Fender, R. P. & Belloni, T. GRS 1915+105 and the disc-jet coupling in accreting black hole systems. *Annu. Rev. Astron. Astrophys.* **42**, 317–364 (2004).
- Belloni, T., Klein-Wolt, M., Méndez, M., van der Klis, M. & van Paradijs, J. A model-independent analysis of the variability of GRS 1915+105. *Astron. Astrophys.* **355**, 271–290 (2000).
- Hynes, R. I. *et al.* The remarkable rapid X-ray, ultraviolet, optical and infrared variability in the black hole XTE J1118+480. *Mon. Not. R. Astron. Soc.* **345**, 292–310 (2003).
- Motch, C., Ilovaisky, S. A. & Chevalier, C. Discovery of fast optical activity in the X-ray source GX 339–4. *Astron. Astrophys.* **109**, L1–L4 (1982).
- Tanaka, Y. & Shibazaki, N. X-ray novae. *Annu. Rev. Astron. Astrophys.* **34**, 607–644 (1996).
- Khargharia, J., Froning, C. S. & Robinson, E. L. Near-infrared spectroscopy of low-mass X-ray binaries: accretion disk contamination and compact object mass determination in V404 Cyg and Cen X-4. *Astrophys. J.* **716**, 1105–1117 (2010).
- Miller-Jones, J. C. A. *et al.* The first accurate parallax distance to a black hole. *Astrophys. J.* **706**, L230–L234 (2009).
- Makino, F. GS 2023+338. *IAU Circ.* **4782** (1989).
- Barthelmy, S. D. *et al.* Swift trigger 643949 is V404 Cyg. *GRB Coord. Netw. Circ.* **17929** (2015).
- Negoro, H. *et al.* MAXI/GSC detection of a new outburst from the Galactic black hole candidate GS 2023+338 (V\* V404 Cyg). *Astron. Telegr.* **7646** (2015).
- Chen, Y. T. *et al.* TAOS early optical observations of V404 Cyg. *Astron. Telegr.* **7722** (2015).
- Golenetskii, S. *et al.* Konus-Wind observation of Galactic transient V404 Cyg in outburst. *GRB Coord. Netw. Circ.* **17938** (2015).
- Uemura, M. *et al.* Rapid optical fluctuations in the black hole binary V4641 Sagittarii. *Publ. Astron. Soc. Jpn* **54**, L79–L82 (2002).
- Zycki, P. T., Done, C. & Smith, D. A. The 1989 May outburst of the soft X-ray transient GS 2023+338 (V404 Cyg). *Mon. Not. R. Astron. Soc.* **309**, 561–575 (1999).
- Belloni, T., Méndez, M., King, A. R., van der Klis, M. & van Paradijs, J. An unstable central disk in the superluminal black hole X-ray binary GRS 1915+105. *Astrophys. J.* **479**, L145–L148 (1997).
- Neilsen, J., Remillard, R. A. & Lee, J. C. The physics of the “heartbeat” state of GRS 1915+105. *Astrophys. J.* **737**, 69–108 (2011).
- Altamirano, D. *et al.* The faint “heartbeats” of IGR J17091–3624: an exceptional black hole candidate. *Astrophys. J.* **742**, L17–L23 (2011).
- Osaki, Y. Dwarf-nova outbursts. *Publ. Astron. Soc. Pacif.* **108**, 39–60 (1996).
- Steehhs, D. *et al.* The not-so-massive black hole in the microquasar GRS 1915+105. *Astrophys. J.* **768**, 185–191 (2013).
- Janiuk, A., Grzedzielski, M., Capitanio, F. & Bianchi, S. Interplay between heartbeat oscillations and wind outflow in microquasar IGR J17091–3624. *Astron. Astrophys.* **574**, A92–A102 (2015).
- Casares, J., Charles, P. A. & Naylor, T. A 6.5-day periodicity in the recurrent nova V404 Cygni implying the presence of a black hole. *Nature* **355**, 614–617 (1992).
- Orosz, J. A. *et al.* A black hole in the superluminal source SAX J1819.3–2525 (V4641 Sgr). *Astrophys. J.* **555**, 489–503 (2001).
- Bagnoli, T. & in’t Zand, J. J. M. Discovery of GRS 1915+105 variability patterns in the rapid burster. *Mon. Not. R. Astron. Soc.* **450**, L52–L56 (2015).
- Hameury, J.-M., Menou, K., Dubus, G., Lasota, J.-P. & Hure, J.-M. Accretion disc outbursts: a new version of an old model. *Mon. Not. R. Astron. Soc.* **298**, 1048–1060 (1998).
- Panopoulou, G., Reig, P. & Blinov, D. Optical polarization of V404 Cyg. *Astron. Telegr.* **7674** (2015).
- Itoh, R. *et al.* Optical and near-infrared polarimetry for V404 Cyg with 1.6m Pirka and 1.5m Kanata telescopes in Japan. *Astron. Telegr.* **7709** (2015).
- Lasota, J.-P. The disc instability model of dwarf novae and low-mass X-ray binary transients. *New Astron. Rev.* **45**, 449–508 (2001).
- Kim, S.-W., Wheeler, J. C. & Mineshige, S. Disk instability and outburst properties of the intermediate polar GK Persei. *Astrophys. J.* **384**, 269–283 (1992).
- King, A. R. & Ritter, H. The light curves of soft X-ray transients. *Mon. Not. R. Astron. Soc.* **293**, L42–L48 (1998).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We acknowledge the variable star observations from the AAVSO International Database contributed by observers worldwide and used in this research. We also thank the INTEGRAL groups for making the products of the ToO data public online at the INTEGRAL Science Data Centre. Work at ASI/A was supported in part by the thematic research program AS-88-TP-A02. A.S.P., E.D.M. and A.A.V. are grateful to the Russian Science Foundation (grant 15-12-30016) for support. R.Ya.I. is grateful for partial support by the grant RUSTAVELI FR/379/6-300/14. We thank H. Maehara, H. Akazawa, K. Hirokawa and J. Lluís for their optical observations. This work was supported by the Grant-in-Aid “Initiative for High-Dimensional Data-Driven Science through Deepening of Sparse Modeling” from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan (25120007 TK and 26400228 YU).

**Author Contributions** M.K. led the campaign, performed optical data analysis and compiled all optical data. K.I. and A.I. performed optical data analysis. T.K., Y.U., D.N. and M.U. contributed to science discussions. S.N., M.S., T.E., T.H. and H.T. performed X-ray data analysis. Other authors than those mentioned above performed optical observations. M.K., K.I., T.K., Y.U., S.N., T.E., M.S. and A.I. wrote the manuscript. T.K., Y.U. and D.N. supervised this project. M.K., K.I., T.K., Y.U., T.E., M.S., D.N., C.L., R.I., M.J.L., F.B.B., D.K., E.P.P., A.S.P., I.E.M., M.R., E.M., W.L.S., S.K., L.M.C., A.I. and M.U. improved the manuscript. All authors have read and approved the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.K. ([mkimura@kustastro.kyoto-u.ac.jp](mailto:mkimura@kustastro.kyoto-u.ac.jp)).



## METHODS

**Detailed methods of optical observations.** Immediately after the detection by Swift/BAT on June 15.77197 UT, the VSNET collaboration team<sup>31</sup> started a world-wide photometric campaign of V404 Cyg. There was also an independent detection by CCD (charge coupled device) photometry on June 16.169 UT<sup>32</sup>. Time-resolved CCD photometry was carried out at 27 sites using 36 telescopes with apertures of dozens of centimetres (Extended Data Table 2). We also used the public AAVSO data<sup>33</sup>. We corrected for bias and flat-fielding in the usual manner, and performed standard aperture photometry. The observers, except for TAOS<sup>34</sup>, used standard filters (B, V, R<sub>C</sub>, I<sub>C</sub>; we write R and I for R<sub>C</sub> and I<sub>C</sub> in the main text and figures for brevity) and measured magnitudes of V404 Cyg relative to local comparison stars whose magnitudes were measured by A. Henden (sequence 15167RN) from the AAVSO Variable Star Database<sup>35</sup>. We applied small zero-point corrections to some observers' measurements. When filtered observations were unavailable, we used unfiltered data to construct the light curve. The exposure times were mostly 2–30 s, with some exceptional cases of 120 s in B band, giving typical time resolution of a few seconds. All of the observation times were converted to BJD.

**Comparison with X-ray observations.** For the Swift/XRT light curves (Fig. 3 and Extended Data Fig. 2), we extracted source events from a region with a 30-pixel radius centred on V404 Cyg. To avoid pile-up effects, we further excluded an inner circular region if the maximum count rate of the XRT raw light curves, binned in 10 s intervals, exceeded 200 counts s<sup>-1</sup>. The inner radii are set to be 10 and 20 pixels at the maximum raw rate of 1,000 counts s<sup>-1</sup> and 2,000 counts s<sup>-1</sup>, respectively, and those for intermediate count rates were determined via linear interpolation between the two points. The presented light curves were corrected for photon losses due to this exclusion by using the *xrtlccorr* tool. In addition, from Fig. 3a, c and d, we can see a time delay in the start of a dip in optical light, relative to that in X-rays. The delay time was ~1 min, which is similar to the reported value of 0–50 s (ref. 36). This was determined by cross-correlating the U-band and X-ray (0.3–10 keV) light curves obtained with Swift/Ultraviolet and Optical Telescope (UVOT) and Swift/XRT on UT 2015 June 21<sup>36</sup>. The observations were carried out when the source showed little rapid optical flickering and no extreme flares, and thus the nature of the lag may be different from that in our observations. We also note that the apparent difference between the Swift/UVOT and the ground-based times<sup>36</sup> is caused by the drift of the clock on board the satellite, to which we have applied the necessary corrections.

**Origin of cyclic dips.** In order to examine the possibility that absorption by gas in the line-of-sight causes the observed violent flux variations in the optical and X-ray bands (Fig. 3), we studied intensity-sliced X-ray spectra. A striking example is shown in Extended Data Fig. 3a. The period shown corresponds to that in Fig. 3a when both the X-ray and optical fluxes exhibited a sudden intensity drop towards the latter part of the period. We divided it into five intervals (T1 to T5; Extended Data Fig. 3a), and generated spectra through the tools *xrtpipeline* and *xrtproducts* in standard pipeline processing. We excluded the central 60-arcsec strip from this Windowed Timing (WT) mode data, to avoid the heavy pile-up effect when the raw count rate exceeds ~150 counts s<sup>-1</sup>. We compared the *νF<sub>ν</sub>* spectra of the five intervals, where the spectra are fitted by a single power-law model multiplied by photoelectric absorption (*phabs* × *pegpwrlw*; in the standard X-ray spectral fitting package XSPEC). The absorbed X-ray flux ranges by two orders of magnitude, from  $2.1 \times 10^{-9}$  ergs s<sup>-1</sup> cm<sup>-2</sup> in T5 to  $3.0 \times 10^{-7}$  ergs s<sup>-1</sup> cm<sup>-2</sup> in T3. However, the best-fit column density and photon index were relatively stable over the five intervals,  $\sim(2-6) \times 10^{-21}$  cm<sup>-2</sup> and  $\sim 1.0-1.5$ , respectively. Since the X-ray spectrum does not show a noticeable rise in column density when the X-ray flux sharply dropped, and since there is no stronger iron edge in the latter part of the observation, absorption cannot be the primary cause of the time variation in our data sets that cover the X-ray and optical bands simultaneously.

**Objects showing violent short-term variations in outburst.** In Extended Data Table 3 we show the list of X-ray binaries that have shown violent short-term variations either in X-rays or in optical wavelengths.

IGR J17091–3624 is known as the second black hole X-ray binary whose X-ray light curves showed a variety of patterns, resembling those of GRS 1915 + 105<sup>18</sup>. The variations observed in the 2011 outburst of this object were classified as  $\rho$  ('heartbeat'),  $\nu$  (similar to class  $\rho$  but with secondary peak after the dips),  $\alpha$  ('rounded-bumps'),  $\beta/\lambda$  (repetitive short-term oscillations after low-quiet period) and  $\mu$  (ref. 18).

The Rapid Burster (RB or MXB 1730–335), a low-mass X-ray binary (LMXB) containing a neutron star (NS), was discovered by Small Astronomy Satellite (SAS-3) observations<sup>37</sup>. This object has been recently reported to show cyclic long X-ray bursts with periods of a few seconds resembling class  $\rho$  ('heartbeat') variations and those with periods of 100–200 s resembling class  $\theta$  ("M"-shaped light curves) variations of GRS 1915 + 105<sup>24</sup>. The emission of the Rapid Burster did not reach the Eddington luminosity during these variations<sup>38</sup>.

V4641 Sgr was originally discovered as a variable star<sup>39</sup> and was long confused with a different variable star, GM Sgr<sup>40</sup>. V4641 Sgr is famous for its short and bright outburst in 1999, which reached a optical magnitude of at least 8.8 mag (refs 41–44). V4641 Sgr showed short-term variations in optical wavelengths during the 2002, 2003 and 2004 outbursts<sup>14,45–47</sup>. It was the first case in which short-term and large-amplitude variations in the optical range during an outburst were detected. V4641 Sgr is classified as a LMXB, and has a long orbital period. Its mass-accretion rate is less than the Eddington rate (except for the 1999 outburst<sup>44,48</sup>). These properties are similar to those of V404 Cyg. However, while the short-term variations of V4641 Sgr seemed to be random, those of V404 Cyg showed repetitive patterns; this is the greatest difference between these two objects. There has been a suggestion that V4641 Sgr is a 'microblazar'<sup>49</sup> because the jets observed during the outburst in 1999 were proposed to have the largest bulk Lorentz factor among known galactic sources<sup>43</sup>.

There are also other X-ray transients showing short-term optical variations (for example, XTE J1118+480 and GX 339–4). However, these two sources are quasi-periodic oscillations (QPOs), characterized by very short periods. The periods are much shorter than those of repetitive patterns (tens of seconds to a few hours) that we discuss in this Letter. Furthermore, the amplitudes of their variations are significantly smaller than those observed in V4641 Sgr<sup>4,50</sup> on timescales longer than tens of seconds.

### Estimation of the disk mass and comparison with the previous outbursts.

Following the method in ref. 15, we estimated the mass stored in the disk at the onset of the outburst. By integrating the X-ray light curve of Swift/BAT and assuming the spectral model C in table 1 in ref. 15, we obtained a value of  $5.0 \times 10^{25}$  g assuming a radiative efficiency of 10% and a distance of  $2.4 \pm 0.2$  kpc (ref. 8). The mass during the 1989 outburst has been updated to  $3.0 \times 10^{25}$  g by using this updated distance. The stored mass in the 2015 outburst was approximately the same as that in the 1989 one. As discussed in ref. 15, these masses are far smaller than the mass of a fully built-up disk, estimated to be  $2.0 \times 10^{28}$  g, if these outbursts were starting at the outermost region.

We compare the published optical light curves of the 1989 and 1938 outbursts<sup>51,52</sup> with our data from the 2015 outburst (Extended Data Fig. 4). We can see that these outbursts have different durations. The 1938 outburst was apparently longer than the others, and it may have had different properties from the 1989 and 2015 ones. The fading rates of the 1989 and 2015 outbursts are significantly larger than those of classical X-ray transients<sup>6</sup>, or of FRED (fast rise and exponential decline)-type outbursts, such as 0.028 mag d<sup>-1</sup> in V518 Per = GRO J0422+32 (ref. 53) and 0.015 mag d<sup>-1</sup> in V616 Mon = A0620–00 (ref. 54). This supports the hypothesis that the outbursts in 1989 and 2015 are different from typical outbursts of classical X-ray transients and that the stored disk mass was a factor of  $\sim 10^3$  smaller in the 1989 and 2015 outbursts than the mass of a fully built up disk.

**Power spectra.** We performed power spectral analyses on BJD 2,457,193, BJD 2,457,196 and BJD 2,457,200. We used the continuous and regularly sampled high-cadence data set obtained by LCO (Extended Data Table 1) with exposure times of 5 s (on BJD 2,457,193) and 2 s (others). The durations of these observations are 1.4, 3.1 and 2.2 h, respectively. Considering the read-out times of 1 s, the Nyquist frequencies of these observations are 0.08 and 0.17 Hz, respectively. The power spectral densities (PSDs) were calculated using *powspec* software in the FTOOLS Xronos package on magnitude measurements. We did not apply de-trending of the light curve since the durations of the individual observations were shorter than the timescale of the global variation of the outburst. The power spectra are well expressed by a power law ( $P \propto f^{-\Gamma}$ ) with an index  $\Gamma$  of  $1.9 \pm 0.1$ ,  $1.8 \pm 0.1$ , and  $2.3 \pm 0.1$  on BJD 2,457,193, 2,457,196 and 2,457,200, respectively (Extended Data Fig. 5). Interpretation of the physical origins on the basis of these variations is difficult, because a power law index of  $\sim 2$  in the PSDs is often observed in natural phenomena. In this region ( $f < 0.01$  Hz), the power originating in the optical variations of V404 Cyg is significantly higher than that of white noise estimated from the observations.

We next summarize the other reports on short-term variations of V404 Cyg during the present outburst. On BJD 2,457,191, this object was observed using the Argos photometer on the 2.1 m Otto Struve Telescope at McDonald Observatory with an exposure time of 2 s<sup>55</sup>. They reported that the power spectrum was dominated by steep red noise. Observations on BJD 2,457,193 and BJD 2,457,194 were also performed using the ULTRACAM attached with the 4.2 m William Herschel Telescope on La Palma observatory with a high time resolution (466.8 ms)<sup>56</sup>. They reported that the variations were dominated by timescales longer than tens of seconds. Although large amplitude flares (0.3–0.4 mag) on timescales shorter than 1 s were reported<sup>57</sup>, these flares may be of different origin. For the variations with timescales longer than 100 s, our results agree with these reports<sup>55,56</sup>.

**Disk radius inferred from final fading rate.** The timescale  $\tau$  of heating/cooling waves in dwarf novae and X-ray transients<sup>58</sup> is a function of the mass of the central object ( $M_1$ ) and radius ( $r$ ) with the form  $\tau \propto \alpha M_1^{-1/2} r^{3/2}$ , where  $\alpha$  is the viscosity

parameter<sup>59</sup>. Here, we estimate the disk radius of V404 Cyg assuming that the timescale of the final fading reflected a dwarf nova-type cooling wave. Using the Kepler data of V344 Lyr and V1504 Cyg, we measured a fading rate of  $1.5 \text{ mag d}^{-1}$  of the normal outbursts immediately preceding superoutbursts. During the outbursts in V344 Lyr and V1504 Cyg<sup>60</sup>, the disk radius is expected to be very close to the 3:1 resonance radius. Adopting a typical mass of a white dwarf in a cataclysmic variable ( $M_1 = 0.83 M_\odot$ ; ref. 61), we estimated the disk radius of V404 Cyg to be  $7.8 \times 10^{10} \text{ cm}$  for a black hole mass of  $9 M_\odot$ . This is much smaller than the radius ( $1.2 \times 10^{12} \text{ cm}$ ) expected for a fully built-up disk<sup>15</sup>.

**SED modelling.** Extended Data Fig. 6a shows the multi-wavelength SED on BJD 2,457,199.431 to 2,457,199.446, when the source was simultaneously observed in the X-ray, ultraviolet (UV) and optical bands. The optical fluxes in the V and I<sub>C</sub> bands are taken from our photometric data averaged over the period. Note that R<sub>C</sub>-band data are also available but not used here, because of the contamination by the continuum strong H $\alpha$  line<sup>62–64</sup>.

The X-ray spectrum is extracted from simultaneous Swift/XRT data (ObsID 00031403058) which were taken in the WT mode. The data are processed through the pipeline processing tool xrt pipeline. The events detected within 20 pixels around the source position are removed to mitigate pile-up effects. The U-band flux is obtained from the Swift/UVOT images with the same ObsID as the XRT, through the standard tool uvot2pha provided by the Swift team. A circular region centred at the source position with a radius of 5 arcsec is adopted as the source extraction region of the UVOT data. The optical, UV and X-ray data are corrected for interstellar extinction/absorption by assuming  $A_V$  (interstellar extinction in the V band) = 4 (ref. 65) and using the extinction curve in ref. 66 and the  $N_H$  (hydrogen column density) versus  $E(B-V)$  relation in ref. 67. Radio data are from the RATAN-600 observation performed in the same period<sup>68</sup>.

The multi-wavelength SED can be reproduced with the disk model<sup>69,70</sup>, which accounts for the emission from the accretion disk, including the effects of Comptonization in the inner disk and reprocessing in the outer disk. We find that partial covering X-ray absorption (using the pcfabs model implemented in the spectral analysis software XSPEC) improves the quality of the fit significantly. The inner-disk temperature is estimated to be  $0.12 \pm 0.01 \text{ keV}$ , and the electron temperature and photon index of the Comptonization component, the ratio between the luminosity of the Compton tail and disk blackbody ( $L_C/L_d$ ), and the fraction of the bolometric flux thermalized in the outer disk ( $f_{\text{out}}$ ), are  $17.5 \pm 0.8 \text{ keV}$ ,  $1.78 \pm 0.03$ ,  $1.17 \pm 0.03$ , and  $1.3^{+0.6}_{-0.8} \times 10^{-2}$ , respectively (the errors in this section represent 90% confidence ranges for one parameter). The inner radius ( $R_{\text{in}}$ ) is estimated to be  $(1.5\text{--}5.4) \times 10^8 \text{ cm}$ , and the outer radius ( $R_{\text{out}}$ ) is  $(2.5 \pm 0.3) \times 10^{12} \text{ cm}$ . The derived value of  $R_{\text{out}}$  is comparable to or even larger than the binary separation ( $\sim 2.2 \times 10^{12} \text{ cm}$ ). However, it could be smaller due to uncertainties in interstellar/circumbinary extinction<sup>71</sup> and/or the contribution of jet emission. For instance, if  $A_V$  is 0.4 mag larger than the assumed value (4.0),  $R_{\text{out}}$  becomes  $(1.9 \pm 0.2) \times 10^{12} \text{ cm}$ . The maximum achievable radius of a stable disk for a  $q$  (mass ratio) = 0.06 object (Extended Data Table 3) is around 0.62A (radius of the 2:1 resonance) to  $\sim 0.7A$  (tidal limit), where  $A$  is the binary separation<sup>72</sup>. Considering the uncertainties, the result of our analysis ( $\gtrsim 0.77A$ ) is compatible with this maximum radius. Our result appears to favour a large  $A_V$  value. For the partial covering absorber, the best-fit value of the column density is  $5.2^{+0.4}_{-0.5} \times 10^{23} \text{ cm}^{-2}$  and that of the covering fraction is  $64 \pm 4\%$ .

The radio SED can be approximated by a power-law with a photon index of  $\sim 1$ , as in other black hole binaries in the low/hard state<sup>73</sup>. This profile is likely to be generated by the optically-thick synchrotron emission from compact jets<sup>74</sup>. Because an optically-thick synchrotron spectrum often extends up to the millimetre to near-infrared bands<sup>75–77</sup>, it may contribute to the optical fluxes, in particular at longer wavelengths. The blackbody emission from the companion, a K3III-type star<sup>7</sup> with a radius of  $\sim 3 R_\odot$  and a temperature of  $\sim 4,320 \text{ K}$ , contributes to the SED negligibly.

Extended Data Figure 6b plots the simultaneous SED on BJD 2,457,191.519 to 2,457,191.524, which is  $\sim 2$  orders of magnitude fainter in the X-ray band than that shown in the left panel. The X-ray, UV and optical data are taken from the Swift data (ObsID 00031403038) and our photometric measurements in the same manner as described above. This SED can be reproduced with the irradiated disk model as well, with somewhat smaller photon index ( $1.43^{+0.02}_{-0.03}$ ) and inner-disk temperature ( $< 0.07 \text{ keV}$ ), and a larger  $f_{\text{out}}$  ( $0.06^{+0.02}_{-0.05}$ ) than those on BJD 2,457,199.431 to 2,457,199.446.

**Time history of the bolometric luminosity.** The bolometric luminosity  $L_{\text{bol}}$  of V404 Cyg is evaluated based on the hard X-rays above  $\sim 15 \text{ keV}$  where the intrinsic spectrum is less affected by an absorption.

We processed the Swift/BAT archival survey data via batsurvey in the HEASoft package to derive count rates with individual exposures of  $\sim 300 \text{ s}$ . Even within this short exposure, photon statistics are good during bright states ( $> 0.05 \text{ counts s}^{-1}$ ). Assuming a Crab-like spectrum ( $1 \text{ Crab} \approx 0.039 \text{ counts s}^{-1}$ ), the BAT count rates  $R$

(counts  $\text{s}^{-1}$ ) are then converted into  $15\text{--}50 \text{ keV}$  flux ( $F_{15-50}$ ) and luminosity ( $L_{15-50}$ ) using  $F_{15-50} = 3.6 \times 10^{-7} R$  (ergs  $\text{s}^{-1} \text{ cm}^{-2}$ ) and a fiducial distance of  $2.4 \text{ kpc}$ , respectively. In Fig. 4, we show  $L_{\text{bol}}$  after multiplying by a conversion factor  $L_{\text{bol}}/L_{15-50} = 7$  determined from SED modelling (previous section). We find that this bolometric correction factor lies within the range 2.5–10 by fitting 19 X-ray(XRT)-optical simultaneous SED in different periods between BJD 2,457,192.019 and 2,457,201.011. Since the BAT survey data are rather sparse, in order to catch shorter-term variations, we further overlaid the INTEGRAL IBIS/ISGRI monitoring in the 25–60 keV band available at ref. 78, assuming a conversion parameter of 1 Crab rate to be  $172.1 \text{ counts s}^{-1}$  and a bolometric correction factor of  $L_{\text{bol}}/L_{25-60} = 9.97$ .

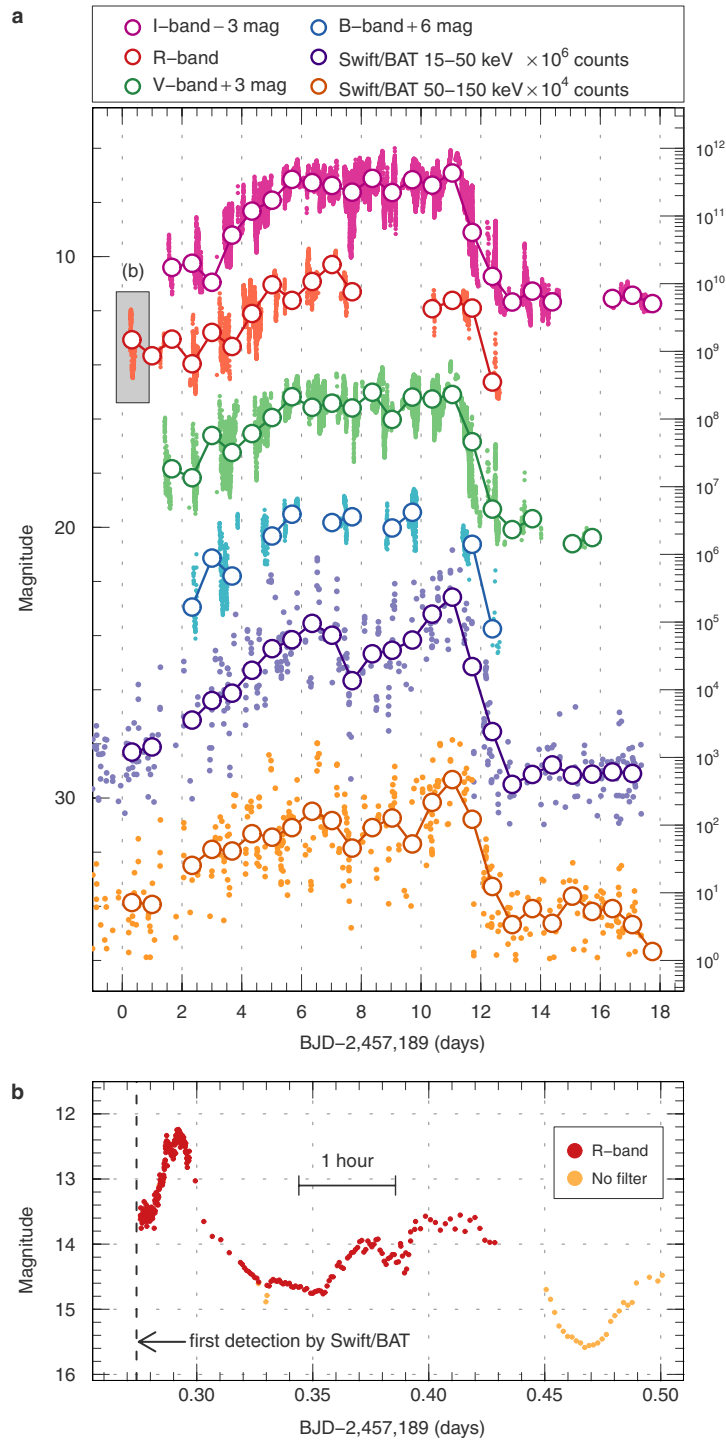
The luminosity was highly variable during the outburst, changing by five orders of magnitude. While V404 Cyg sometimes reaches the Eddington luminosity ( $L_{\text{Edd}}$ ) at the peak of multiple sporadic flares, it also repeatedly dropped below  $\sim 10\%$  of  $L_{\text{Edd}}$  (Fig. 4). At earlier phases of this outburst, the characteristic oscillation already occurred during a lower luminosity state, as discussed in the main text.

**Sample size.** No statistical methods were used to predetermine sample size.

- Kato, T. *et al.* Variable Star Network: world center for transient object astronomy and variable stars. *Publ. Astron. Soc. Jpn* **56**, S1–S54 (2004).
- Muyllaert, E. V404 Cyg going into outburst? *BAAVSS Alert* 4101 (2015); <https://groups.yahoo.com/neo/groups/baavss-alert/conversations/messages/4101>.
- AAVSO American Association of Variable Star Observers. Download data. <http://www.aavso.org/data-download/> (accessed 4 July 2015).
- Lehner, M. J. *et al.* The Taiwanese-American Occultation Survey: the multi-telescope robotic observatory. *Publ. Astron. Soc. Pacif.* **121**, 138–152 (2009).
- AAVSO American Association of Variable Star Observers. Variable star plotter. <http://www.aavso.org/vsp> (accessed 4 July 2015).
- Gandhi, P. *et al.* Correlated optical and X-ray variability in V404 Cyg. *Astron. Teleg.* **7727** (2015).
- Lewin, W. H. G. *et al.* The discovery of rapidly repetitive X-ray bursts from a new source in Scorpius. *Astrophys. J.* **207**, L95–L99 (1976).
- Bagnoli, T., in't Zand, J. J. M., Galloway, D. K. & Watts, A. L. Indications for a slow rotator in the Rapid Burster from its thermonuclear bursting behaviour. *Mon. Not. R. Astron. Soc.* **431**, 1947–1955 (2013).
- Goranskij, V. P. Variable stars in Sagittarius. *Astronomicheskii Tsirkulyar* **1024**, 3–4 (1978).
- Samus, N. N. *et al.* V4641 Sagittarii and GM Sagittarii. *IAU Circ.* **7277** (1999).
- Stubbings, R. *et al.* GM Sagittarii and SAX J1819.3–2525 = XTE J1819–254. *IAU Circ.* **7253** (1999).
- Kato, T., Uemura, M., Stubbings, R., Watanabe, T. & Monard, B. Preoutburst activity of V4641 Sgr = SAX J1819.3–2525: possible existence of 2.5-day period. *Inform. Bull. Variable Stars* **4777** (1999).
- Hjellming, R. M. *et al.* Light curves and radio structure of the 1999 September transient event in V4641 Sagittarii (=XTE J1819–254 = SAX J1819.3–2525). *Astrophys. J.* **544**, 977–992 (2000).
- Uemura, M. *et al.* The 1999 optical outburst of the fast X-ray nova, V4641 Sagittarii. *Publ. Astron. Soc. Jpn* **54**, 95–101 (2002).
- Uemura, M. *et al.* Outburst and post-outburst active phase of the black hole X-ray binary, V4641 Sgr in 2002. *Publ. Astron. Soc. Jpn* **56**, S61–S75 (2004).
- Uemura, M. *et al.* Optical observation of the 2003 outburst of a black hole X-ray binary, V4641 Sagittarii. *Publ. Astron. Soc. Jpn* **56**, 823–829 (2004).
- Uemura, M. *et al.* Outburst of a black hole X-ray binary V4641 Sgr in 2004 July. *Inform. Bull. Variable Stars* **5626**, 1–4 (2005).
- Revnivtsev, M., Sunyaev, R., Gilfanov, M. & Churazov, E. V4641 Sgr — a super-Eddington source enshrouded by an extended envelope. *Astron. Astrophys.* **385**, 904–908 (2002).
- Mirabel, I. F. & Rodríguez, L. F. Sources of relativistic jets in the Galaxy. *Annu. Rev. Astron. Astrophys.* **37**, 409–443 (1999).
- Imamura, J. N., Kristian, J., Middleditch, J. & Steiman-Cameron, T. Y. The 8 second optical quasi-periodic oscillations in GX 339–4. *Astron. Astrophys.* **365**, 312–316 (1990).
- Casares, J., Charles, P. A., Jones, D. H. P., Rutten, R. G. M. & Callanan, P. J. Optical studies of V404 Cyg, the X-ray transient GS 2023+338. I — the 1989 outburst and decline. *Mon. Not. R. Astron. Soc.* **250**, 712–725 (1991).
- Wagner, R. M. *et al.* Optical identification of the X-ray source GS 2023+338 as V404 Cygni. *Astrophys. J.* **378**, 293–297 (1991).
- Chevalier, C. & Ilovaisky, S. A. CCD photometry of GRO J0422+32 during activity and quiescence. *Astron. Astrophys.* **297**, 103–114 (1995).
- Whelan, J. A. J. *et al.* Spectroscopic observations of the X-ray nova A0620–00. *Mon. Not. R. Astron. Soc.* **180**, 657–673 (1977).
- Hynes, R. I., Robinson, E. L. & Morales, J. Rapid optical photometry of V404 Cyg. *Astron. Teleg.* **7677** (2015).
- Gandhi, P. *et al.* Sub-second multi-band optical timing of V404 Cyg with ULTRACAM. *Astron. Teleg.* **7686** (2015).
- Hynes, R. I., Robinson, E. L. & Morales, J. Further rapid optical photometry of V404 Cyg. *Astron. Teleg.* **7710** (2015).
- Cannizzo, J. K. On the relative rates of decay of the optical and soft X-ray fluxes in dwarf nova outbursts. *Astrophys. J.* **473**, L41–L44 (1996).
- Meyer, F. Transition waves in accretion disks. *Astron. Astrophys.* **131**, 303–308 (1984).

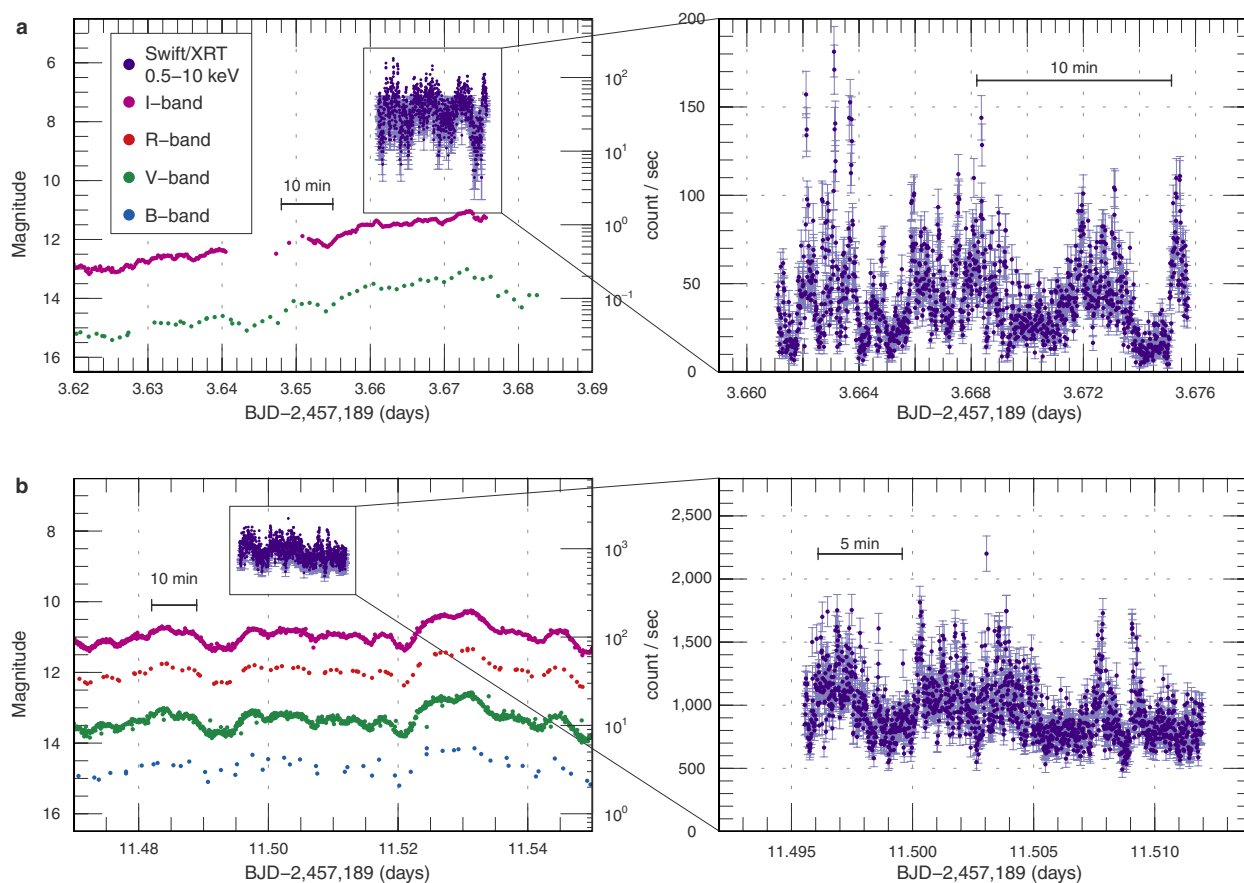
60. Cannizzo, J. K., Smale, A. P., Wood, M. A., Still, M. D. & Howell, S. B. The Kepler light curves of V1504 Cygni and V344 Lyrae: a study of the outburst properties. *Astrophys. J.* **747**, 117–128 (2012).
61. Savoury, C. D. *et al.* Cataclysmic variables below the period gap: mass determinations of 14 eclipsing systems. *Mon. Not. R. Astron. Soc.* **415**, 2025–2041 (2011).
62. Munoz-Darias, T., Sanchez, D. M. & Casares, J. Optical spectroscopy of V404 Cyg: evolution of the P Cygni profiles. *Astron. Telegr.* **7669** (2015).
63. Caballero-Garcia, M. D., Castro-Tirado, A. J., Oates, S. & Jeong, S. Early optical spectroscopy follow-up of V404 Cyg with GTC/OSIRIS. *Astron. Telegr.* **7699** (2015).
64. Scarpaci, J., Maitra, D., Hynes, R. & Markoff, S. Multi-band optical observations of V404 Cygni and correlated spectral changes. *Astron. Telegr.* **7737** (2015).
65. Casares, J., Charles, P. A., Naylor, T. & Pavlenko, E. P. Optical studies of V404 Cygni the X-ray transient GS 2023+338 – part three – the secondary star and accretion disc. *Mon. Not. R. Astron. Soc.* **265**, 834–852 (1993).
66. Cardelli, J. A., Clayton, G. C. & Mathis, J. S. The relationship between infrared, optical, and ultraviolet extinction. *Astrophys. J.* **345**, 245–256 (1989).
67. Bohlin, R. C., Savage, B. D. & Drake, J. F. A survey of interstellar H I from  $\Lambda\alpha$  absorption measurements. II. *Astrophys. J.* **224**, 132–142 (1978).
68. Trushkin, S. A., Nizhelskij, N. A. & Tybulev, P. G. The inverted radio spectrum of the flare in V404 Cyg. *Astron. Telegr.* **7667** (2015).
69. Gies, D. R. *et al.* Stellar wind variations during the X-ray high and low states of Cygnus X-1. *Astrophys. J.* **678**, 1237–1247 (2008).
70. Kaptein, R. G. *et al.* Discovery of 1RXS J171824.2–402934 as an X-ray burster. *Astron. Astrophys.* **358**, L71–L74 (2000).
71. Hynes, R. I. *et al.* The quiescent spectral energy distribution of V404 Cyg. *Mon. Not. R. Astron. Soc.* **399**, 2239–2248 (2009).
72. Osaki, Y. & Meyer, F. Early humps in WZ Sge stars. *Astron. Astrophys.* **383**, 574–579 (2002).
73. Fender, R. P. Powerful jets from black hole X-ray binaries in low/hard X-ray states. *Mon. Not. R. Astron. Soc.* **322**, 31–42 (2001).
74. Blandford, R. D. & Königl, A. Relativistic jets as compact radio sources. *Astrophys. J.* **232**, 34–48 (1979).
75. Corbel, S. & Fender, R. P. Near-infrared synchrotron emission from the compact jet of GX 339–4. *Astrophys. J.* **573**, L35–L39 (2002).
76. Gandhi, P. *et al.* A variable mid-infrared synchrotron break associated with the compact jet in GX 339–4. *Astrophys. J.* **740**, L13–L19 (2011).
77. Russell, T. D. *et al.* The accretion-ejection coupling in the black hole candidate X-ray binary MAXI J1836–194. *Mon. Not. R. Astron. Soc.* **439**, 1390–1402 (2014).
78. INTEGRAL Science Data Centre. INTEGRAL data analysis. <http://www.isdc.unige.ch/integral/analysis#QLAources> (8 August 2015).
79. Wang, J. H. *et al.* Early optical brightening in GRB 071010B. *Astrophys. J.* **679**, L5–L8 (2008).
80. Kloppenborg, B. K., Pieri, R., Eggenstein, H.-B., Maravelias, G. & Pearson, T. A demonstration of accurate wide-field V-band photometry using a consumer-grade DSLR camera. *J. Am. Assoc. Variable Star Obs.* **40**, 815–833 (2012).
81. Alcock, C. *et al.* TAOS: The Taiwanese-American Occultation Survey. *Earth Moon Planets* **92**, 459–464 (2003).
82. Zhang, Z.-W. *et al.* The TAOS project: results from seven years of survey data. *Astron. J.* **146**, 14–23 (2013).
83. Casares, J. & Charles, P. A. Optical studies of V404 Cyg, the X-ray transient GS 2023+338. IV. the rotation speed of the companion star. *Mon. Not. R. Astron. Soc.* **271**, L5–L9 (1994).
84. Wijnands, R., Yang, Y. J. & Altamirano, D. The enigmatic black hole candidate and X-ray transient IGR J17091–3624 in its quiescent state as seen with XMM-Newton. *Mon. Not. R. Astron. Soc.* **422**, L91–L95 (2012).
85. Reid, M. J. *et al.* A parallax distance to the microquasar GRS 1915+105 and a revised estimate of its black hole mass. *Astrophys. J.* **796**, 2–9 (2014).
86. Iyer, N., Nandi, A. & Mandal, S. Determination of the mass of IGR J17091–3624 from “spectro-temporal” variations during the onset phase of the 2011 outburst. *Astrophys. J.* **807**, 108–116 (2015).
87. Sala, G. *et al.* Constraints on the mass and radius of the accreting neutron star in the Rapid Burster. *Astrophys. J.* **752**, 158–164 (2012).
88. MacDonald, R. K. D. *et al.* The black hole binary V4641 Sagittarii: activity in quiescence and improved mass determinations. *Astrophys. J.* **784**, 2–20 (2014).
89. Fender, R. P. *et al.* MERLIN observations of relativistic ejections from GRS 1915+105. *Mon. Not. R. Astron. Soc.* **304**, 865–876 (1999).
90. King, A. L. *et al.* An extreme X-ray disk wind in the black hole candidate IGR J17091–3624. *Astrophys. J.* **746**, L20–L24 (2012).
91. Szkody, P. *et al.* V404 Cygni. *IAU Circ.* **4794** (1989).



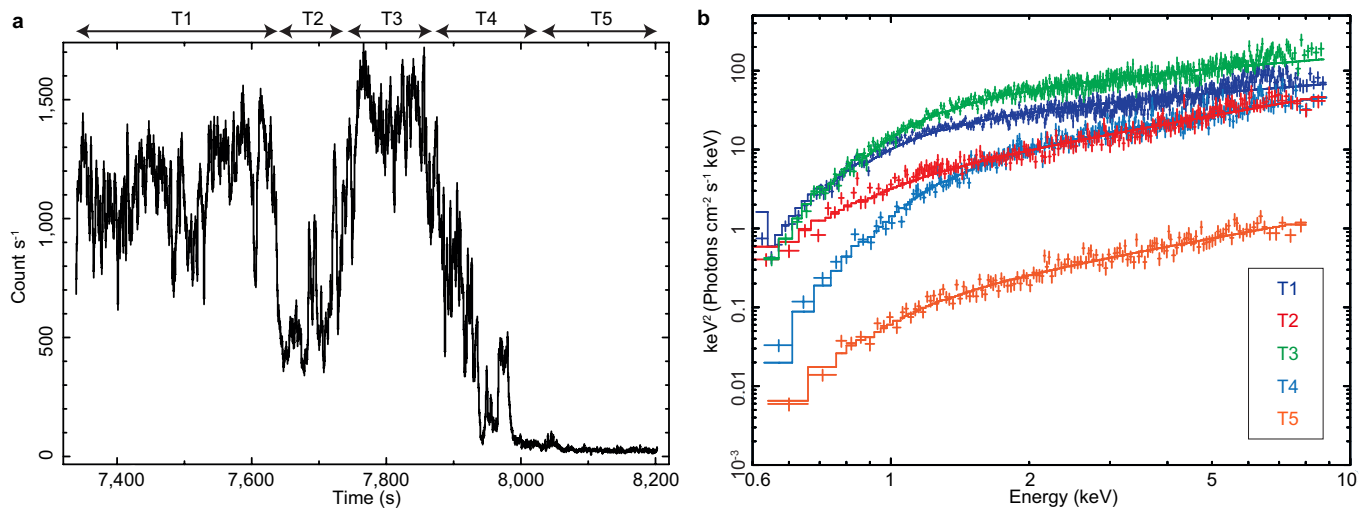


**Extended Data Figure 1 | Optical and X-ray light curves of V404 Cyg during an outburst in 2015 June–July. a**, Overall multi-colour light curves and Swift/BAT light curves. The plotted points are averaged for every 0.67 days. **b**, An enlarged view of the shaded box in **a** (the first detection of short-term variations). On BJD 2,457,203, the mean

magnitude dropped below  $V = 17.0$ . Superimposed on this rapid fading, the amplitude of variations became progressively smaller and smaller. After BJD 2,457,205, the mean magnitude seemed to be constant, and the outburst virtually ended.



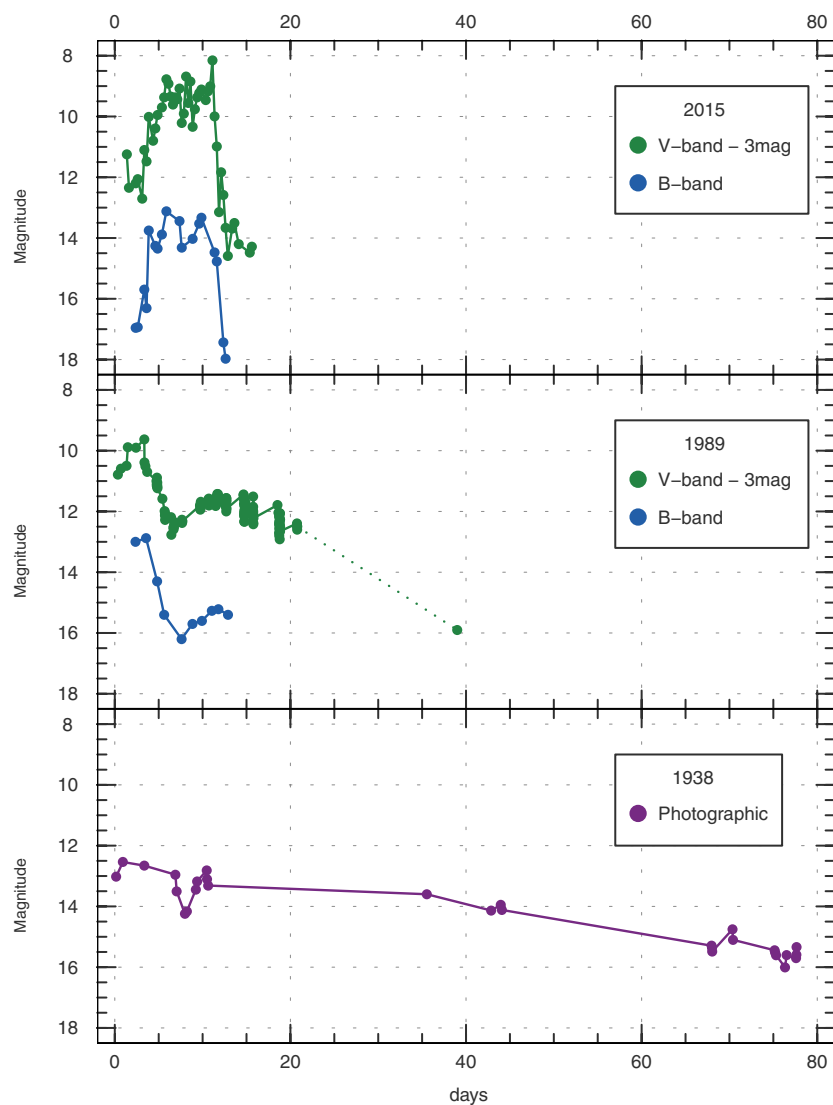
**Extended Data Figure 2 | Additional examples of simultaneous optical and X-ray observations of V404 Cyg in the 2015 outburst.** Data shown in Fig. 3 are excluded. **a, b**, Main panels, correlations on BJD 2,457,192 (a) and BJD 2,457,200 (b); right panels, Swift/XRT light curves on linear scales. Navy blue error bars,  $\pm 1\sigma$ .



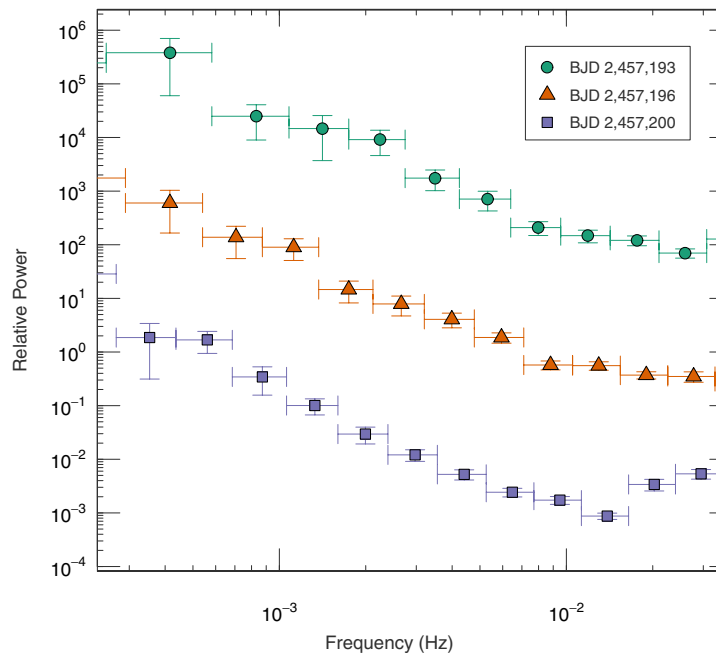
**Extended Data Figure 3 | Example of the soft X-ray light curve and spectra during the dip-type oscillation in the 2015 outburst of V404 Cyg.** **a**, The  $\sim 860$ -s-long Swift/XRT raw light curve (BJD 2,457,194.125–2,457,194.135, ObsID 00031403040) without pile-up correction, same as

the X-ray data in Fig. 3a. **b**, Time-sliced soft X-ray spectra with pile-up correction, in the intervals of T1 to T5 determined in **a**. The exposures of individual spectra are  $\sim 100$ – $300$  s. Error bars,  $\pm 1\sigma$ .



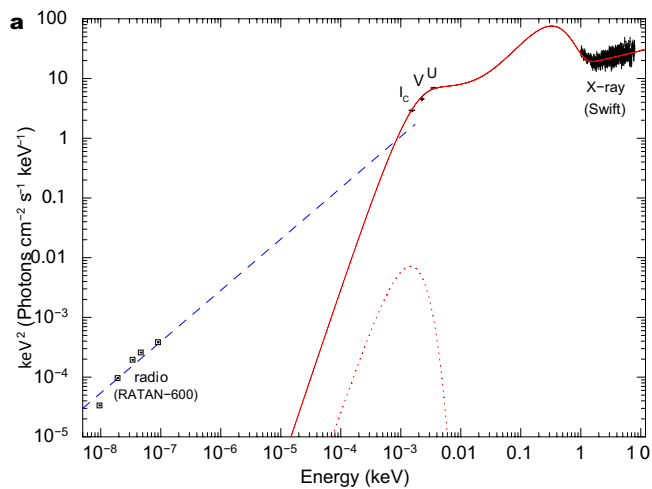


**Extended Data Figure 4 | Comparison of the 1938, 1989 and 2015 outbursts of V404 Cyg.** The horizontal axis represents days BJD  $- 2,429,186$ , BJD  $- 2,447,673$  and BJD  $- 2,457,189$ , respectively. Photographic magnitudes are approximately the same as B band.

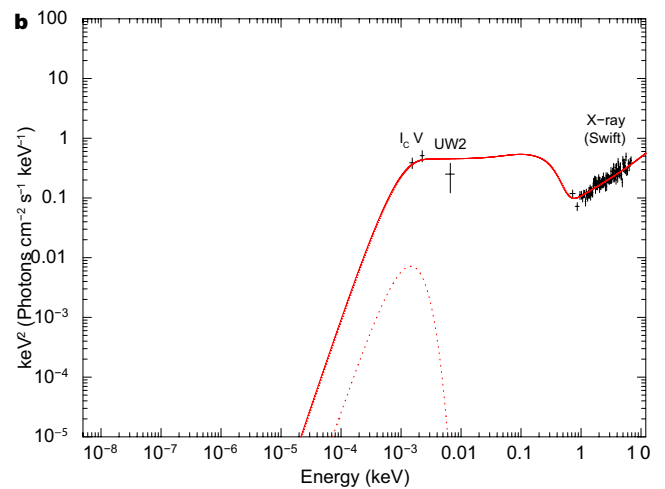


**Extended Data Figure 5 | Power spectral densities of the early stage, the middle stage, and the later stage in the 2015 outburst of V404 Cyg.** Power spectral densities of the fluctuations on BJD 2,457,193 (top, circles), BJD 2,457,196 (middle, triangles) and BJD 2,457,200 (bottom, rectangles). The abscissa and ordinate denote the frequency in

Hz and the power in arbitrary units, respectively. For better visualization, the obtained spectrum is multiplied by  $8 \times 10^{-4}$  on BJD 2,457,196 and by  $10^{-4}$  on BJD 2,457,200.  $\pm 1\sigma$  error bars obtained from relevant  $\chi^2$  distributions of the power spectra.



**Extended Data Figure 6 | Simultaneous, extinction-corrected multi-wavelength SEDs of V404 Cyg.** **a, b,** The intervals shown are BJD 2,457,199.431–2,457,199.446 (**a**) and BJD 2,457,191.519–2,457,191.524 (**b**). The optical (V and  $I_c$ ) fluxes are averaged over the intervals; error bars, s.e. The X-ray, U- and UW2-band data are obtained with Swift; error bars,  $\pm 1\sigma$ . The radio fluxes (open squares) are compiled from the



RATAN-600 results at BJD 2,457,199.433 (ref. 68). The red solid and dotted lines show the contribution of emissions from the irradiated disk with Comptonization and from the companion star, respectively. The blue dashed line approximates the radio SED, which is extended to the optical bands for illustrative purposes.

Extended Data Table 1 | A log of photometric observations of the 2015 outburst of V404 Cyg

Start*	End*	Mag <sup>†</sup>	Error <sup>‡</sup>	N <sup>§</sup>	Obs	Band <sup>¶</sup>	Start*	End*	Mag <sup>†</sup>	Error <sup>‡</sup>	N <sup>§</sup>	Obs	Band <sup>¶</sup>
0.274	0.295	13.24	0.032	215	TAO	R	7.314	7.511	12.54	0.065	67	CRI	V
0.282	0.499	15.34	0.101	37	PZN	CR	7.314	7.512	11.45	0.060	66	CRI	R <sub>c</sub>
0.386	0.426	15.18	0.036	86	PZN	R <sub>c</sub>	7.315	7.511	10.36	0.061	65	CRI	I <sub>c</sub>
0.386	0.426	14.31	0.040	20	CRI&PZN	R <sub>c</sub>	7.422	7.588	10.52	0.011	1501	IMI	I <sub>c</sub>
1.137	1.192	14.92	0.024	61	PZN	CR	7.427	7.670	11.35	0.035	1104	deM	I <sub>c</sub>
1.274	1.398	13.93	0.086	20	PZN	R <sub>c</sub>	7.675	7.802	10.81	0.012	1961	LCO	I <sub>c</sub>
1.283	1.284	11.40	0.000	2	KW2	I <sub>c</sub>	7.707	7.945	10.70	0.019	1003	SWI	I <sub>c</sub>
1.283	1.284	12.50	0.000	2	KW2	V	7.744	7.907	12.98	0.036	350	GFB	V
1.283	1.284	11.95	0.068	2	KW2	B	8.030	8.300	10.22	0.020	535	KU1	I <sub>c</sub>
1.551	1.670	13.48	0.029	191	deM	I <sub>c</sub>	8.032	8.035	11.69	0.036	5	OKU	V
1.627	1.810	14.98	0.009	2430	LCO	CR	8.038	8.297	10.27	0.016	1022	OKU	I <sub>c</sub>
2.109	2.517	15.07	0.024	224	PZN	CR	8.038	8.128	12.03	0.040	81	loh	CR
2.277	2.404	14.57	0.104	35	PZN	R <sub>c</sub>	8.152	8.214	12.41	0.028	103	Wnm	cG
2.341	2.522	14.19	0.044	231	DPV	CR	8.360	8.543	9.95	0.015	68	CRI	I <sub>c</sub>
2.354	2.529	15.41	0.056	158	DPV	V	8.394	8.619	10.41	0.011	623	Kai	I <sub>c</sub>
2.354	2.529	14.12	0.049	158	DPV	R <sub>c</sub>	8.419	8.671	10.17	0.012	1129	deM	I <sub>c</sub>
2.380	2.505	17.19	0.075	61	Ter	B	8.709	8.859	13.42	0.036	413	RIT	V
2.380	2.505	15.52	0.072	61	Ter	V	8.969	9.043	10.87	0.019	296	Sac	I <sub>c</sub>
2.381	2.506	14.33	0.062	61	Ter	R <sub>c</sub>	8.993	9.154	10.55	0.024	608	Kis	I <sub>c</sub>
2.406	2.524	14.65	0.024	354	Ter	CR	9.006	9.044	12.77	0.032	40	Sac	V
2.422	2.615	14.43	0.045	151	Kai	I <sub>c</sub>	9.179	9.315	12.49	0.053	146	PZN	CR
2.423	2.609	14.43	0.045	147	Kai	R <sub>c</sub>	9.224	9.229	12.59	0.149	5	OKU	V
2.446	2.669	13.46	0.021	667	deM	I <sub>c</sub>	9.239	9.300	10.84	0.020	152	OKU	I <sub>c</sub>
2.742	2.859	13.91	0.009	2652	LCO	CR	9.382	9.620	10.40	0.002	643	Kai	I <sub>c</sub>
3.801	3.341	12.55	0.048	1216	TAO	R	9.414	9.595	10.24	0.003	428	NDJ	I <sub>c</sub>
3.251	3.524	16.10	0.054	186	Ter	B	9.577	9.841	13.54	0.020	620	RIT	B
3.252	3.525	14.45	0.051	183	Ter	V	9.607	9.798	12.21	0.005	4709	LCO	V
3.252	3.524	13.41	0.044	177	Ter	R <sub>c</sub>	9.635	9.828	10.13	0.010	1823	LCO	I <sub>c</sub>
3.260	3.529	13.58	0.017	1278	Ter	CR	9.744	9.911	12.38	0.031	350	GFB	V
3.266	3.308	13.54	0.086	48	PZN	CR	10.027	10.028	11.90	0.018	3	Kis	V
3.271	3.307	13.64	0.091	40	PZN	R <sub>c</sub>	10.029	10.201	10.54	0.011	837	Kis	I <sub>c</sub>
3.410	3.489	15.80	0.095	38	CRI	B	10.387	10.619	10.46	0.020	611	Kai	I <sub>c</sub>
3.411	3.488	14.36	0.071	37	CRI	V	10.415	10.670	10.38	0.013	1389	deM	I <sub>c</sub>
3.411	3.488	13.17	0.062	37	CRI	R <sub>c</sub>	10.744	10.910	11.99	0.010	349	GFB	V
3.411	3.489	12.01	0.058	37	CRI	I <sub>c</sub>	11.182	11.300	9.41	0.012	99	KU1	I <sub>c</sub>
3.419	3.588	14.48	0.048	189	RPc	V	11.291	11.298	10.55	0.003	112	TAO	R
3.428	3.553	14.52	0.056	128	Trt	V	11.339	11.514	10.51	0.018	406	DPV	I <sub>c</sub>
3.430	3.519	12.25	0.023	597	IMI	I <sub>c</sub>	11.348	11.554	13.10	0.014	730	Trt	V
3.435	3.673	12.47	0.020	1036	deM	I <sub>c</sub>	11.372	11.515	13.15	0.019	335	DPV	V
3.525	3.650	12.64	0.075	37	COO	I <sub>c</sub>	11.385	11.592	11.00	0.015	490	Kai	I <sub>c</sub>
3.530	3.820	14.53	0.076	165	Kis	V	11.421	11.673	11.32	0.021	1314	deM	I <sub>c</sub>
3.819	3.821	10.39	0.014	2	Kis	I <sub>c</sub>	11.460	11.624	13.71	0.097	70	JSa	V
3.998	4.057	11.49	0.038	149	KU1	I <sub>c</sub>	11.483	11.603	13.53	0.016	374	RJV	V
4.059	4.311	12.04	0.036	397	Mdy	R <sub>c</sub>	11.590	11.679	14.43	0.014	730	LCO	V
4.187	4.316	11.88	0.022	169	TAO	R	11.679	11.834	12.95	0.008	3859	LCO	I <sub>c</sub>
4.435	4.673	11.66	0.021	1089	deM	I <sub>c</sub>	12.228	12.232	15.49	0.139	5	OKU	V
4.546	4.649	13.41	0.041	82	Kis	V	12.234	12.271	12.95	0.028	177	TAO	R
4.579	4.637	11.32	0.008	1416	LCO	I <sub>c</sub>	12.302	12.334	13.81	0.011	311	TAO	R
4.976	4.978	12.14	0.034	5	Kis	V	12.386	12.611	13.89	0.025	484	Kai	I <sub>c</sub>
4.979	4.981	10.04	0.004	3	Kis	I <sub>c</sub>	12.405	12.670	14.08	0.022	640	deM	I <sub>c</sub>
5.070	5.223	11.19	0.042	254	Mdy	R <sub>c</sub>	12.484	12.599	16.87	0.031	237	RJV	V
5.426	5.481	12.75	0.054	36	CRI	B	13.058	13.314	15.81	0.013	211	Mdy	R <sub>c</sub>
5.427	5.481	13.96	0.057	36	CRI	V	13.199	13.334	15.97	0.048	1772	TAO	R
5.427	5.480	11.66	0.048	35	CRI	R <sub>c</sub>	13.382	13.594	14.28	0.014	467	Kai	I <sub>c</sub>
5.427	5.480	10.57	0.044	36	CRI	I <sub>c</sub>	13.415	13.670	14.08	0.022	640	deM	I <sub>c</sub>
5.448	5.633	10.45	0.010	840	deM	I <sub>c</sub>	13.438	13.473	13.91	0.040	93	NDJ	I <sub>c</sub>
5.595	5.670	10.25	0.020	25	COO	I <sub>c</sub>	14.014	14.021	17.21	0.075	5	OKU	V
5.724	5.954	9.85	0.007	920	SWI	I <sub>c</sub>	14.026	14.168	14.75	0.022	97	OKU	I <sub>c</sub>
5.745	5.911	11.69	0.011	346	GFB	V	14.043	14.276	14.70	0.016	361	Kis	I <sub>c</sub>
5.923	5.949	10.32	0.013	24	COO	I <sub>c</sub>	14.379	14.499	17.00	0.030	52	Trt	CV
6.011	6.015	12.51	0.016	5	OKU	V	14.421	14.565	14.85	0.012	152	RPc	I <sub>c</sub>
6.019	6.076	10.27	0.005	154	OKU	V	14.422	14.614	14.56	0.007	248	NDJ	I <sub>c</sub>
6.146	6.157	10.01	0.050	4	KW2	I <sub>c</sub>	14.504	14.517	17.15	0.114	5	Trt	V
6.146	6.157	12.02	0.121	4	KW2	V	14.601	14.810	15.56	0.003	1830	LCO	CR
6.146	6.157	13.15	0.159	2	KW2	B	15.166	15.276	16.92	0.234	664	TAO	R
6.182	6.281	10.55	0.048	129	Aka	R <sub>c</sub>	15.356	15.549	16.04	0.007	244	DPV	CR
6.210	6.280	12.41	0.060	64	Aka	V	15.364	15.550	17.41	0.024	42	DPV	V
6.293	6.554	11.31	0.030	85	CRI	R <sub>c</sub>	15.434	15.559	14.84	0.017	81	RPc	I <sub>c</sub>
6.295	6.550	12.30	0.037	83	CRI	V	15.694	15.762	14.45	0.008	166	SWI	I <sub>c</sub>
6.346	6.428	11.84	0.028	93	PZN	R <sub>c</sub>	16.092	16.142	14.60	0.397	5	TAO	R
3.356	6.543	9.94	0.011	412	DPV	I <sub>c</sub>	16.302	16.377	16.07	0.013	26	PZN	CR
6.363	6.521	12.25	0.010	572	Trt	V	16.320	16.525	14.44	0.010	129	CRI	I <sub>c</sub>
6.369	6.406	10.09	0.008	334	DPV	I <sub>c</sub>	16.344	16.435	14.50	0.012	52	DPV	I <sub>c</sub>
6.430	6.615	12.38	0.022	418	RJV	V	16.516	16.530	14.44	0.030	6	RPc	I <sub>c</sub>
6.584	6.827	12.65	0.005	5910	LCO	V	16.680	16.937	14.36	0.006	335	SWI	I <sub>c</sub>
6.592	6.861	10.54	0.012	794	RIT	I <sub>c</sub>	17.358	17.518	14.48	0.006	218	DPV	I <sub>c</sub>
6.717	6.944	10.36	0.007	942	SWI	I <sub>c</sub>	17.418	17.671	14.78	0.006	309	deM	I <sub>c</sub>
6.745	6.912	12.35	0.010	347	GFB	V	17.440	17.575	14.62	0.014	43	RPc	I <sub>c</sub>
6.919	6.950	10.35	0.076	24	COO	I <sub>c</sub>	18.297	18.336	17.37	0.254	470	TAO	R
7.056	7.057	13.27	0.010	3	Kis	V	19.328	19.332	16.45	0.258	68	TAO	R
7.057	7.137	10.55	0.019	295	Kis	V	19.403	19.451	14.82	0.011	33	DPV	I <sub>c</sub>
7.115	7.147	10.01	0.018	45	Aka	R <sub>c</sub>	19.423	19.498	14.79	0.026	17	RPc	I <sub>c</sub>
7.144	7.150	9.87	0.016	18	KW2	I <sub>c</sub>	19.712	19.761	16.61	0.008	60	GFB	CV
7.144	7.150	11.76	0.068	18	KW2	V	20.435	20.592	14.98	0.008	90	RPc	I <sub>c</sub>
7.144	7.150	13.44	0.184	2	KW2	B	21.023	21.031	15.34	0.012	10	RPc	I <sub>c</sub>
7.313	7.512	13.79	0.063	67	CRI	B							

Start and end dates of observations, mean magnitudes,  $1\sigma$  of mean magnitudes, numbers of observations, observers' codes, and filters are summarized. Note that observers for TAOS used custom made filters close to the union of standard R and V<sup>34,79</sup>, but the magnitude reported in the present Letter was approximately calibrated to standard R.

\*JD - 2,457,189 (days).

†Mean magnitude.

‡ $1\sigma$  of mean magnitude.

§Number of observations.

|| Observer's code: PZN (IKI GRB follow up network), CRI (Crimean Observatory Team), deM (E. de Miguel), DPV (P. A. Dubovsky), Ter (Terskol Observatory), Kai (K. Kasai), NDJ (N. James), RPc (R. D. Pickard), Trt (T. Tordai), COO (L. Cook), Kis (S. Kiyota), KU1 (Kyoto Univ. Team), Mdy (Y. Maeda), LCO (C. Littlefield), RIT (M. Richmond), RJV (R. Javier), GFB (W. Goff), SWI (W. L. Stein), OKU (Osaka Kyoiku Univ. team), Sac (A. Miyashita), IMI (I. Miller), TAO (TAOS Team), KW2 (H. Maehara), Aka (H. Akazawa), Wnm (K. Hirose) and JSa (J. Lluís).

¶Filter: B, V, R<sub>c</sub>, I<sub>c</sub> are the standard Johnson-Cousins system. 'CR' and 'CV' mean unfiltered CCD photometry with zero point adjustment in R and V, respectively. 'cG' means green (G) channel output in a digital single-lens reflex camera, which gives an approximate response close to V (ref. 80).



Extended Data Table 2 | List of instruments for optical observations

CODE	Telescope (& CCD)	Observatory (or Observer)	Site
PZN	1m Zeiss-1000 Tien Shan +Apogee Alta	Astronomical Observatory	Almaty, Kazakhstan
	40cm ORI-40+FLI ML09000	ISON-Khureltogot	Mongolia
	70cm+FLI AS-32+FLI IMG6303E	Abastumani observatory	Georgia
CRI	1.25m AZT-11+FLI ProLine PL230	Crimean astrophysical observatory	Crimea
	38cm K-380+Apogee E47	Crimean astrophysical observatory	Crimea
deM	35cm SC+QSI-516wsg	Observatorio Astronomico del CIECEM	Huelva, Spain
DPV	28cm SC+MII G2-1600	Astronomical Observatory on Kolonica	Slovakia
	35cm SC+MII G2-1600	Astronomical Observatory on Kolonica	Slovakia
	VNT 1m+FLI PL1001E	Astronomical Observatory on Kolonica	Slovakia
Ter	Zeiss-600 60cm+SBIG STL-1001E	Terskol Observatory	Russia
	S2C 35cm	Terskol Observatory	Russia
Kai	28cm SC+ST7XME	Kiyoshi Kasai	Switzerland
NDJ	28cm SC+ST9XE	Nick James	UK
RPc	FTN 2.0m+E2V 42-40	LCOGT*	Hawaii, USA
	35cmSC+SXV-H9 CCD	Roger D. Pickard	UK
Trt	25cm ALCCD5.2 (QHY6)	Tamás Tordai	Budapest, Hungary
COO	T07 <sup>†</sup> 43cm+STL-1100M	AstroCamp Observatory	Nerpio, Spain
	T21 <sup>†</sup> 43cm+FLI-PL6303E	iTelescope.Net Mayhill	New Mexico, USA
	T11 <sup>†</sup> 50cm+FLI ProLine PL11002M	iTelescope.Net Mayhill	New Mexico, USA
Kis	25cm SC+Alta F47	Seiichiro Kiyota	Kamagaya, Japan
	T18 <sup>†</sup> 32cm+STXL-6303E	AstroCamp Observatory	Nerpio, Spain
	T5 <sup>†</sup> 25cm+ST-10XME	iTelescope.Net Mayhill	New Mexico, USA
	T24 <sup>†</sup> 61cm+FLI-PL09000	Sierra Remote Observatoy	California, USA
KU1	40cm SC+ST-9XEI	Kyoto U. Team	Kyoto, Japan
Mdy	35cm SC+ST10XME	Yutaka Maeda	Nagasaki, Japan
LCO	60cm+Apogee Alta U42 CCD	Van Vleck Observatory	Connecticut, USA
	40cm+SBIG STL-6303	Van Vleck Observatory	Connecticut, USA
RIT	30cm+ST-9E	RIT Observatory	New York, USA
RJV	LX200R 40cm+ST8 XME	Observatorio de Cantabria	Spain
GFB	CDK 50cm+Apogee U6	William Goff	California, USA
SWI	C14 35cmSC+ST10XME	William L. Stein	New Mexico, USA
OKU	51cm+Andor DW936N-BV	OKU Astronomical Observatory	Osaka, Japan
Sac	20cmL+ST-7XMEi	Atsushi Miyashita	Tokyo, Japan
IMi	35cm SC+SXVR-H16	Furzehill Observatory	UK
TAO	TAOS-B <sup>‡</sup> 50cm+SI800 E2V47-20	Lulin Observatory	Taiwan
	TAOS-D <sup>‡</sup> 50cm+SI800 E2V47-20	Lulin Observatory	Taiwan

Observers' codes (see Extended Data Table 1), names of telescopes and CCD cameras, observatory (or observer) and sites are summarized.

\*Las Cumbres Observatory Global Telescope Network.

<sup>†</sup>iTelescope.net.

<sup>‡</sup>The Taiwanese-American Occultation Survey (TAOS)<sup>34,81,82</sup>.

Extended Data Table 3 | Basic information on objects showing violent short-term variations in outbursts

	V404 Cyg	GRS 1915+105	IGR J17091-3624	Rapid Burster	V4641 Sgr
<b>Orbital period [d]</b>	6.47129(7) (ref. 83)	33.85(16) (ref. 20)	>4 (ref. 84)	–	2.81678 (ref. 23)
<b>Compact object</b>	BH	BH	BH	NS	BH
<b>Spectrum of the secondary</b>	K3III (ref. 7)	K–M (ref. 20)	–	–	B9III (ref. 23)
<b><math>M_1</math> ( <math>M_\odot</math> )</b>	9.0(0.6) (ref. 7)	10.1(0.6) (ref. 85)	11.8–13.7 (ref. 86)	1.1(0.3) (ref. 87)	7.1(0.3) (ref. 88)
<b><math>q = M_2 / M_1</math> (Mass ratio)</b>	0.06 (ref. 7)	0.042(0.024) (ref. 20)	–	–	0.45(0.05) (ref. 88)
<b><math>i</math> [deg] (Inclination angle)</b>	67(3) (ref. 7)	66(2) (ref. 89)	50–70 (ref. 90)	–	72.3(4.1) (ref. 88)
<b>V magnitude minimum</b>	18.4 (ref. 91)	–	–	–	13.8 (ref. 88)
<b>V magnitude maximum</b>	10.9 (This work)	–	–	–	8.8 (ref. 41)

Shown are orbital period, nature of the compact object, spectrum of the secondary, mass of the central object ( $M_1$ ), mass ratio ( $q$ ), inclination angle ( $i$ ), minimum magnitude (V band), and maximum magnitude (V band) on V404 Cyg, GRS 1915+105, IGR 17091–3624, the Rapid Burster, and V4641 Sgr.  $M_2$ , mass of the secondary star. References are cited as follows: 7, 20, 23, 41, 83–91.

# A continuum from clear to cloudy hot-Jupiter exoplanets without primordial water depletion

David K. Sing<sup>1</sup>, Jonathan J. Fortney<sup>2</sup>, Nikolay Nikolov<sup>1</sup>, Hannah R. Wakeford<sup>1</sup>, Tiffany Kataria<sup>1</sup>, Thomas M. Evans<sup>1</sup>, Suzanne Aigrain<sup>3</sup>, Gilda E. Ballester<sup>4</sup>, Adam S. Burrows<sup>5</sup>, Drake Deming<sup>6</sup>, Jean-Michel Désert<sup>7</sup>, Neale P. Gibson<sup>8</sup>, Gregory W. Henry<sup>9</sup>, Catherine M. Huitson<sup>7</sup>, Heather A. Knutson<sup>10</sup>, Alain Lecavelier des Etangs<sup>11</sup>, Frederic Pont<sup>1</sup>, Adam P. Showman<sup>4</sup>, Alfred Vidal-Madjar<sup>11</sup>, Michael H. Williamson<sup>9</sup> & Paul A. Wilson<sup>11</sup>

Thousands of transiting exoplanets have been discovered, but spectral analysis of their atmospheres has so far been dominated by a small number of exoplanets and data spanning relatively narrow wavelength ranges (such as 1.1–1.7 micrometres). Recent studies show that some hot-Jupiter exoplanets have much weaker water absorption features in their near-infrared spectra than predicted<sup>1–5</sup>. The low amplitude of water signatures could be explained by very low water abundances<sup>6–8</sup>, which may be a sign that water was depleted in the protoplanetary disk at the planet's formation location<sup>9</sup>, but it is unclear whether this level of depletion can actually occur. Alternatively, these weak signals could be the result of obscuration by clouds or hazes<sup>1–4</sup>, as found in some optical spectra<sup>3,4,10,11</sup>. Here we report results from a comparative study of ten hot Jupiters covering the wavelength range 0.3–5 micrometres, which allows us to resolve both the optical scattering and infrared molecular absorption spectroscopically. Our results reveal a diverse group of hot Jupiters that exhibit a continuum from clear to cloudy atmospheres. We find that the difference between the planetary radius measured at optical and infrared wavelengths is an effective metric for distinguishing different atmosphere types. The difference correlates with the spectral strength of water, so that strong water absorption lines are seen in clear-atmosphere planets and the weakest features are associated with clouds and hazes. This result strongly suggests that primordial water depletion during formation is unlikely and that clouds and hazes are the cause of weaker spectral signatures.

We observed the transits of eight hot Jupiters as part of a spectral survey of exoplanet atmospheres with the Hubble Space Telescope (HST). The eight planets covered in our survey (WASP-6b, WASP-12b, WASP-17b, WASP-19b, WASP-31b, WASP-39b, HAT-P-1b and HAT-P-12b) span a large range of planetary temperature, surface gravity, mass and radii, allowing for an exploration of hot-Jupiter atmospheres across a broad range of physical parameters (see Table 1). In this survey, we observed all eight planets in the full optical wavelength range (0.3–1.01  $\mu\text{m}$ ) using the Space Telescope Imaging Spectrograph (STIS) instrument. We also used the Wide Field Camera 3 (WFC3) instrument to observe transits of WASP-31b and HAT-P-1b in the near-infrared (1.1–1.7  $\mu\text{m}$ ), and used additional WFC3 programmes to observe transits of four other survey targets (WASP-12b, WASP-17b, WASP-19b and HAT-P-12b). The HST survey was complemented by photometric transit observations of all eight targets at 3.6  $\mu\text{m}$  and 4.5  $\mu\text{m}$  using the Spitzer Space Telescope Infrared Array Camera (IRAC) instrument. We analysed the survey targets in conjunction with HST and Spitzer data from the two best-studied hot Jupiters to

date, HD 209458b (ref. 1) and HD 189733b (ref. 5), giving a total of ten exoplanets in our comparative study with transmission spectra between 0.3  $\mu\text{m}$  and 5  $\mu\text{m}$  (see Extended Data Table 1 for a detailed list of the observations).

Our data reduction methods followed those in our previous studies<sup>3,4,11–14</sup>, in which the transmission spectra of WASP-19b, WASP-12b, HAT-P-1b, WASP-6b and WASP-31b were presented (see Methods for further details). The transit light curves<sup>15</sup> of the band-integrated spectra were fitted simultaneously with detector systematics, with all HST and Spitzer transit data used to determine the planets' orbital system parameters (inclination, stellar density and transit ephemeris), which were then fixed to the weighted mean values in the subsequent analysis measuring the transmission spectra. To create the broadband transmission spectrum, we extracted various wavelength bins for the HST STIS and WFC3 spectra and separately fitted each bin for the planet-to-star radius ratio  $R_p/R_*$  and detector systematics. The uncertainties for each data point were rescaled, based on the standard deviation of the residuals, and any systematic errors correlated in time were measured using the binned residuals<sup>16</sup>.

The resulting transmission spectra are shown in Fig. 1 and exhibit a variety of spectral absorption features due to Na, K and H<sub>2</sub>O, as well as strong optical scattering slopes (for example, WASP-6b and HAT-P-12b). Planets such as WASP-39b show prominent alkali absorption lines with pressure-broadened wings, whereas other planets such as WASP-31b show strong but narrow alkali features, which implies that these planets are limited to lower atmospheric pressures. H<sub>2</sub>O vapour has been predicted to be an important source of opacity for hot-Jupiter atmospheres<sup>17–19</sup>, and it is detected in five of the eight exoplanets where WFC3 spectra are available<sup>1–5,13,14</sup>. However, the amplitude of the H<sub>2</sub>O absorption varies greatly across the ten planets, ranging from features that are very pronounced (as in WASP-19b)<sup>14</sup> to those that are much smaller than expected (HD 209458b)<sup>1</sup> or even absent (WASP-31b)<sup>4</sup>.

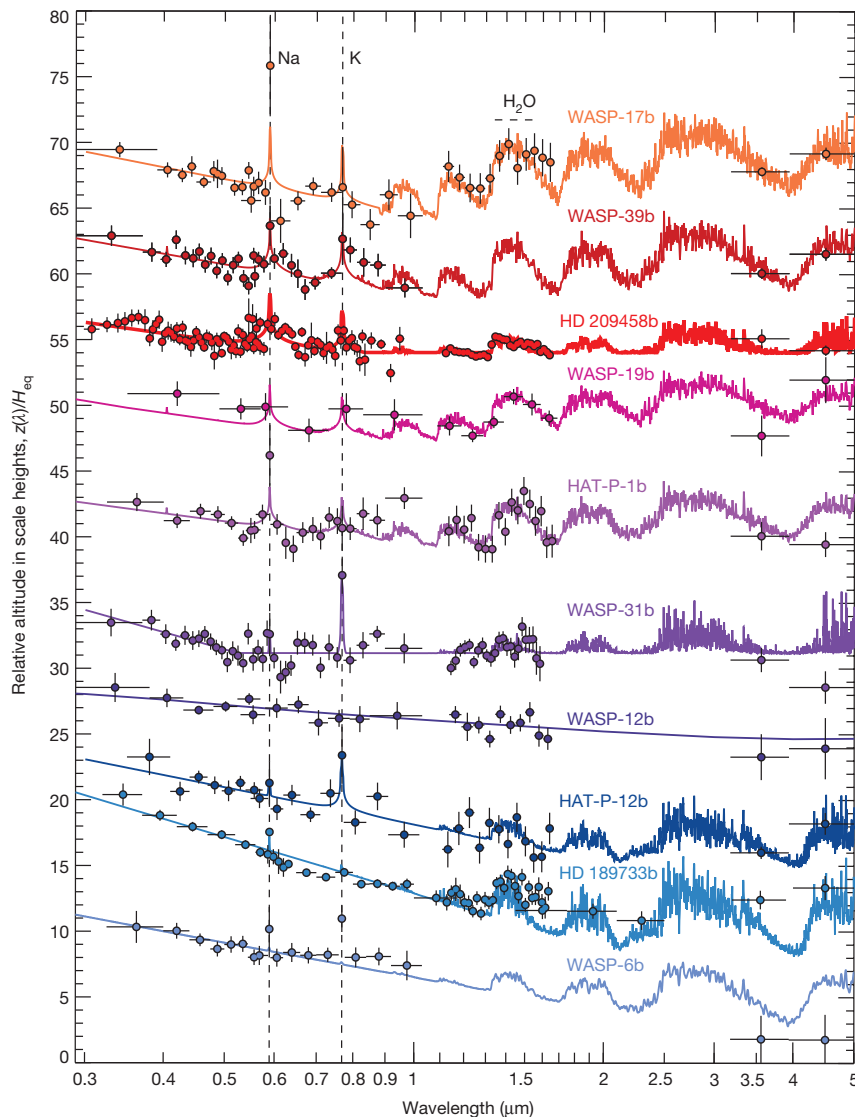
Previous studies using HST/WFC3 spectra have shown that HD 209458b, HD 189733b and WASP-12b have low-amplitude water features<sup>1,3,5</sup>, which can be attributed to a severe depletion of atmospheric H<sub>2</sub>O abundance relative to solar values<sup>6–8</sup>. Any such depletion would be a remnant of planet formation, as H<sub>2</sub>O is expected to be well mixed in a hot atmosphere, such that currently measured molecular abundances would be consistent with primordial values. The depletion of water vapour can occur beyond a protoplanetary disk's snow line<sup>9</sup>, where water is found predominantly as solid ice. Therefore, a hot Jupiter with a large depletion in H<sub>2</sub>O gas would imply that the planet formed at large orbital distances beyond the snow line and,

<sup>1</sup>Astrophysics Group, School of Physics, University of Exeter, Stocker Road, Exeter EX4 4QL, UK. <sup>2</sup>Department of Astronomy and Astrophysics, University of California, Santa Cruz, California 95064, USA. <sup>3</sup>Department of Physics, University of Oxford, Keble Road, Oxford OX1 3RH, UK. <sup>4</sup>Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA. <sup>5</sup>Department of Astrophysical Sciences, Peyton Hall, Princeton University, Princeton, New Jersey 08544, USA. <sup>6</sup>Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA. <sup>7</sup>Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, Colorado 80309, USA. <sup>8</sup>European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748 Garching bei München, Germany. <sup>9</sup>Center of Excellence in Information Systems, Tennessee State University, Nashville, Tennessee 37209, USA. <sup>10</sup>Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA. <sup>11</sup>CNRS, Institut d'Astrophysique de Paris, UMR 7095, 98 bis boulevard Arago, 75014 Paris, France.

**Table 1 | Physical parameters of hot Jupiters and associated spectral results**

Name	$T_{\text{eq}}$ (K)	$g$ (m s <sup>-2</sup> )	$R_p$ ( $R_J$ )	$M_p$ ( $M_J$ )	$P$ (days)	$\log R'_{\text{HK}}$	$\Delta Z_{\text{UB-LM}}/H_{\text{eq}}$	$\Delta Z_{\text{J-LM}}/H_{\text{eq}}$	H <sub>2</sub> O amplitude (%)	Features	Reference
WASP-17b	1,740	3.6	1.89	0.51	3.73	-5.531	$-0.80 \pm 0.36$	$-1.48 \pm 0.71$	$94 \pm 29$	Na, H <sub>2</sub> O	
WASP-39b	1,120	4.1	1.27	0.28	4.06	-4.994	$0.10 \pm 0.41$			Na, K	
HD 209458b	1,450	9.4	1.36	0.69	3.52	-4.970	$0.73 \pm 0.36$	$-0.49 \pm 0.36$	$32 \pm 5$	Aer, Na, H <sub>2</sub> O	
WASP-19b	2,050	14.2	1.41	1.14	0.79	-4.660	$1.04 \pm 1.79$	$-1.97 \pm 1.32$	$105 \pm 20$	H <sub>2</sub> O	14
HAT-P-1b	1,320	7.5	1.32	0.53	4.46	-4.984	$2.01 \pm 0.81$	$0.19 \pm 0.93$	$68 \pm 19$	Na, H <sub>2</sub> O	12, 13
WASP-31b	1,580	4.6	1.55	0.48	3.40	-5.225	$2.15 \pm 0.77$	$1.25 \pm 0.77$	$31 \pm 12$	Aer, K	4
WASP-12b	2,510	11.6	1.73	1.40	1.09	-5.500	$3.76 \pm 1.59$	$1.65 \pm 1.47$	$38 \pm 34$	Aer	3
HAT-P-12b	960	5.6	0.96	0.21	3.21	-5.104	$4.14 \pm 0.77$	$1.37 \pm 0.79$	$17 \pm 23$	Aer, K	
HD 189733b	1,200	21.4	1.14	1.14	2.22	-4.501	$5.52 \pm 0.50$	$-0.56 \pm 0.50$	$53.6 \pm 9.6$	Aer, Na, H <sub>2</sub> O	5, 10
WASP-6b	1,150	8.7	1.22	0.50	3.36	-4.741	$8.49 \pm 1.33$			Aer, K	11

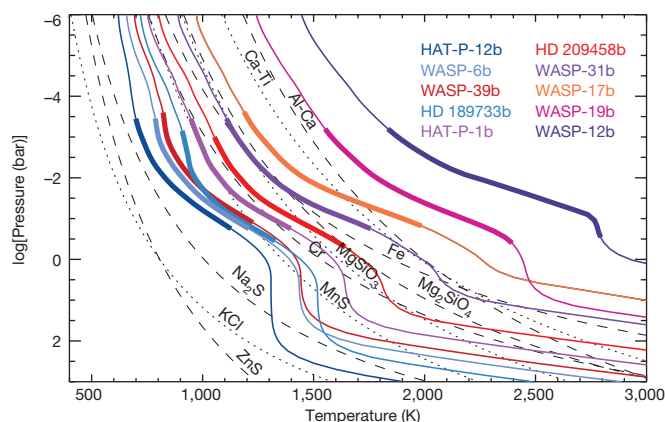
The listed physical parameters are based on data compiled from our HST and Spitzer results<sup>3,4,10-14</sup> and online databases. Sources for published spectral results are also listed. Atmospheric features detected of cloud or haze aerosols, sodium, potassium and water are listed (Aer, Na, K and H<sub>2</sub>O, respectively). The equilibrium temperature  $T_{\text{eq}}$  assumes zero albedo and uniform redistribution. Also listed are the surface gravity,  $g$ ; radius of the planet,  $R_p$ ; planet mass,  $M_p$ ; orbital period,  $P$ ; and Ca II H and K stellar activity index  $\log R'_{\text{HK}}$ .  $R_J$  is the radius of Jupiter and  $M_J$  is the mass of Jupiter.  $\Delta Z_{\text{UB-LM}}/H_{\text{eq}}$  gives the difference in pressure scale heights between the optical and mid-infrared transmission spectra, while  $\Delta Z_{\text{J-LM}}/H_{\text{eq}}$  is the difference between the near- and mid-infrared (see Methods). The atmospheric scale height,  $H_{\text{eq}} = kT_{\text{eq}}/(\mu g)$ , is estimated using the planet-specific equilibrium temperature and assuming a H/He atmosphere with a mean molecular mass of  $\mu = 2.3$  atomic mass units. The H<sub>2</sub>O amplitude is measured using the WFC3 data, taking the average radii from 1.34  $\mu\text{m}$  to 1.49  $\mu\text{m}$  and subtracting it from the average value between 1.22  $\mu\text{m}$  to 1.33  $\mu\text{m}$ , then dividing that value by the theoretical difference as calculated by models<sup>16</sup> assuming clear atmospheres and solar abundances.



**Figure 1 | HST/Spitzer transmission spectral sequence of hot-Jupiter survey targets.** Solid coloured lines show fitted atmospheric models with prominent spectral features indicated. The spectra have been offset, ordered by values of  $\Delta Z_{\text{UB-LM}}$  (the altitude difference between the blue-optical and mid-infrared; Table 1). Horizontal and vertical error bars indicate the wavelength spectral bin and  $1\sigma$  measurement uncertainties,

respectively. Planets with predominantly clear atmospheres (top) show prominent alkali and H<sub>2</sub>O absorption, with infrared radii values commensurate with or higher than the optical altitudes. Very hazy and cloudy planets (bottom) have strong optical scattering slopes, narrow alkali lines and H<sub>2</sub>O absorption that is partially or completely obscured.



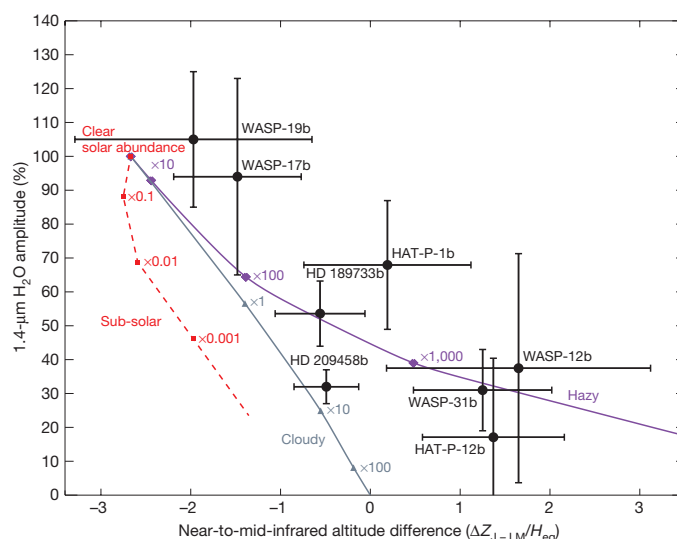


**Figure 2 | Pressure-temperature profiles and condensation curves.** Profiles are calculated for each planet from one-dimensional non-grey radiative transfer models<sup>17</sup>, which assume planet-wide average conditions in chemical equilibrium at solar abundances, and clear atmospheres. Profiles take into account incident stellar fluxes as well as the planetary interior fluxes that are appropriate given each planet's known mass and radius. Dashed and dotted lines are calculations of condensation curves of chemical species expected to condense in planetary and brown dwarf atmospheres<sup>25</sup>. The thicker portions of the pressure-temperature profiles indicate the pressures probed in transmission.

during its inward orbital migration, avoided accretion and dissolution of icy planetesimals as well as the subsequent accretion of an appreciable amount of H<sub>2</sub>O-rich gas. Such scenarios have been proposed for Jupiter<sup>20,21</sup> based on Galileo probe measurements<sup>22</sup> that indicate it to be a water-poor gas giant, although the measurements were affected by local meteorology<sup>22</sup>.

However, it is possible that these weak water absorption bands could be attributed to cloud opacity, which have yielded featureless transmission spectra for a number of transiting exoplanets<sup>23,24</sup>. For simplicity, we define a cloud as a grey opacity source, and a haze as one that yields a Rayleigh-scattering-like opacity, which could be due to small (sub-micrometre size) particles. Silicate or higher-temperature cloud condensates are expected to dominate the hotter atmospheres, like those observed for brown dwarfs, while in cooler atmospheres sulfur-bearing compounds are expected to play an important part in the condensation chemistry<sup>25,26</sup>. In Fig. 2, we plot model atmospheric pressure-temperature profiles for the planets in our comparative study and compare them to the condensation curves for the expected cloud-forming molecules. The base, or bottom, of a condensate cloud is expected to form where the planetary pressure-temperature profiles cross the condensation curve; in this case, Cr, MnS, MgSiO<sub>3</sub>, Mg<sub>2</sub>SiO<sub>4</sub> and Fe are possible condensates. For example, the spectra of WASP-31b shows clouds<sup>4</sup>, which probably form at pressures of about 10 mbar and can be explained by Fe or MgSiO<sub>4</sub> condensates. However, the curves alone cannot explain cloudy versus cloud-free planets, because hazy planets such as HAT-P-12b and WASP-12b do not cross condensation curves at observable pressures. Therefore, atmospheric circulation must also play a part, as vertical mixing allows for particles to be lofted and maintained at pressures probed in transmission at the terminators. Additionally, equatorial eastward superrotation arising from day-night temperature variations can allow clouds that form on the nightside to be transported to the terminator<sup>27</sup>.

We compare spectral features from our large survey to both analytic<sup>4,26</sup> and radiative-transfer models assuming varying degrees of clouds and hazes<sup>17,18</sup>. In order to evaluate the spectral behaviour of the sample as a whole, we define and measure three broadband spectral indices, which can then be compared to both the observational data and the theoretical models (see Table 1 and Methods). We first define an index  $\Delta Z_{UB-LM}$  that compares the relative strength of scattering, which is strongest at blue-optical (0.3–0.57  $\mu\text{m}$ ) wavelengths, to that



**Figure 3 | Transmission spectral index diagram of  $\Delta Z_{j-LM}$  versus H<sub>2</sub>O amplitude.** Black points show the altitude difference between the near-infrared and the mid-infrared spectral features ( $\Delta Z_{j-LM}$ ) versus the amplitude of the 1.4- $\mu\text{m}$  H<sub>2</sub>O feature for eight of ten targets (see Table 1). Error bars represent the  $1\sigma$  measurement uncertainties. Purple and grey lines show model trends for hazy and cloud atmospheres, respectively, with increasing Rayleigh scattering haze and grey cloud deck opacity corresponding to 10 $\times$ , 100 $\times$  and 1,000 $\times$  the solar value. We also show clear-atmosphere models with sub-solar abundances of 0.1 $\times$ , 0.01 $\times$  and 0.001 $\times$  the solar value (red line). WASP-6b and WASP-39b are not included because there are currently no HST WFC3 data for these two planets.

of molecular absorption, which is strongest at mid-infrared (3–5  $\mu\text{m}$ ) wavelengths and dominated by H<sub>2</sub>O, CO and CH<sub>4</sub>. We also define  $\Delta Z_{j-LM}$  to measure the relative strength between the near-infrared continuum (1.22–1.33  $\mu\text{m}$ , located between strong H<sub>2</sub>O absorption bandheads) and the mid-infrared molecular absorption. Lastly, we quantify the amplitude of the H<sub>2</sub>O absorption feature seen in the WFC3 data, calculating the ratio of the observed feature to that of radiative transfer models<sup>17</sup> assuming clear atmospheres and solar abundances.

Comparisons between these indices (Fig. 3, Extended Data Figs 1 and 2) show trends between cloudy and cloud-free planets. When comparing the  $\Delta Z_{j-LM}$  index to the H<sub>2</sub>O amplitude (Fig. 3), the hot-Jupiter transmission spectra strongly favour models in which the H<sub>2</sub>O amplitude is lower owing to obscuration by hazes and clouds, rather than to lower abundances (5.9 $\sigma$  significance). Contaminating effects of persistent unocculted star spots<sup>5</sup> and plages<sup>28</sup> have been proposed in order to mimic the optical haze-scattering signature of hot Jupiters (particularly HD 189733b, which orbits an active star; see Methods). However, our survey sample is sufficiently varied in stellar activity, such that we find no correlation between stellar activity and the strength of the optical scattering slope (as measured by the  $\Delta Z_{UB-LM}$  index) for planets in our sample (Extended Data Fig. 3). One of the main distinguishing features between hazy atmospheres and those that are clear and have sub-solar abundances resides in the near-infrared continuum, measured with the WFC3 spectra. The presence of haze raises the level of the near-infrared continuum relative to the mid-infrared continuum, leading to high  $\Delta Z_{j-LM}$  index values with low near-infrared H<sub>2</sub>O amplitudes (Extended Data Fig. 4). In clear-atmosphere models, where the abundances are lower, the continuum level drops at both near- and mid-infrared wavelengths, accompanied by a reduction in the amplitude of absorption features, resulting in  $\Delta Z_{j-LM}$  index values that are too low to explain the data (Fig. 3).

The hot-Jupiter transmission spectra ordered by the  $\Delta Z_{UB-LM}$  spectral index reveals a continuum from clear atmospheres to atmospheres with strong clouds and hazes (Fig. 1 and Table 1). The presence of clouds has also been inferred for brown dwarf atmospheres, which

have similar temperatures to hot Jupiters<sup>29</sup>. Although there is a well defined sequence from the warmer, cloudy L-dwarfs to the cooler, clear T-dwarfs<sup>29,30</sup>, hot Jupiters do not exhibit a strong relationship of temperature to cloud formation, given that both cloudy and not cloudy planets appear throughout the entire 1,000–2,500 K temperature range (Fig. 2).

We suggest that the difference between hot Jupiters and brown dwarfs is due to the vertical temperature structure of hot-Jupiter atmospheres. Hot Jupiters have very much steeper pressure–temperature profiles compared to isolated brown dwarfs, owing to the strong incident stellar flux heating the top of the planetary atmosphere (see Extended Data Fig. 5). Since cloud condensation curves run nearly parallel to hot-Jupiter profiles, a relatively small temperature shift (about 100 K) could easily move a cloud base by a factor of tens or hundreds in pressure, in or out of the visible atmosphere. In comparison, because brown dwarfs have shallow pressure–temperature profiles, clouds will form in the visible atmosphere across a very wide temperature range. Furthermore, the expected nearly isothermal region of a hot-Jupiter profile at pressures from about 1 bar to 100 bar may cause some planets, but not others, to have cloud materials cold-trapped at depth, out of the visible atmosphere. Given this temperature sensitivity, the role of clouds in hot Jupiters may appear almost stochastic from planet to planet. In addition, hot Jupiters have a wider range of gravities and metallicities, both of which will affect the planet's atmospheric temperature structure, circulation and condensate formation.

Future studies will benefit greatly from broad atmospheric surveys that can further distinguish between clear and cloudy exoplanets. If the  $\Delta Z_{\text{UB-LM}}$ ,  $\Delta Z_{\text{J-LM}}$  and  $\text{H}_2\text{O}$  indices can be measured in advance of such surveys, planets with clear atmospheres can be identified and studied in greater detail, allowing reliable chemical abundances to be measured and thus providing valuable constraints on formation models.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 June; accepted 1 October 2015.

Published online 14 December 2015.

- Deming, D. *et al.* Infrared transmission spectroscopy of the exoplanets HD 209458b and XO-1b using the Wide Field Camera-3 on the *Hubble Space Telescope*. *Astrophys. J.* **774**, 95 (2013).
- Line, M. R. *et al.* A near-infrared transmission spectrum for the warm Saturn HAT-P-12b. *Astrophys. J.* **778**, 183 (2013).
- Sing, D. K. *et al.* HST hot-Jupiter transmission spectral survey: evidence for aerosols and lack of TiO in the atmosphere of WASP-12b. *Mon. Not. R. Astron. Soc.* **436**, 2956–2973 (2013).
- Sing, D. K. *et al.* HST hot-Jupiter transmission spectral survey: detection of potassium in WASP-31b along with a cloud deck and Rayleigh scattering. *Mon. Not. R. Astron. Soc.* **446**, 2428–2443 (2015).
- McCullough, P. R. *et al.* Water vapor in the spectrum of the extrasolar planet HD 189733b. I. The transit. *Astrophys. J.* **791**, A55 (2014).
- Madhusudhan, N. *et al.*  $\text{H}_2\text{O}$  abundances in the atmospheres of three hot Jupiters. *Astrophys. J.* **791**, L9 (2014).
- Madhusudhan, N. *et al.* Toward chemical constraints on hot Jupiter migration. *Astrophys. J.* **794**, L12 (2014).
- Seager, S. *et al.* On the dayside thermal emission of hot Jupiters. *Astrophys. J.* **632**, 1122–1131 (2005).
- Öberg, K. I., Murray-Clay, R. & Bergin, E. A. The effects of snowlines on C/O in planetary atmospheres. *Astrophys. J.* **743**, L16 (2011).
- Pont, F. *et al.* The prevalence of dust on the exoplanet HD 189733b from *Hubble* and *Spitzer* observations. *Mon. Not. R. Astron. Soc.* **432**, 2917–2944 (2013).
- Nikolov, N. *et al.* HST hot-Jupiter transmission spectral survey: haze in the atmosphere of WASP-6b. *Mon. Not. R. Astron. Soc.* **447**, 463–478 (2015).

- Nikolov, N. *et al.* *Hubble Space Telescope* hot Jupiter transmission spectral survey: a detection of Na and strong optical absorption in HAT-P-1b. *Mon. Not. R. Astron. Soc.* **437**, 46–66 (2014).
- Wakeford, H. R. *et al.* HST hot Jupiter transmission spectral survey: detection of water in HAT-P-1b from WFC3 near-IR spatial scan observations. *Mon. Not. R. Astron. Soc.* **435**, 3481–3493 (2013).
- Huitson, C. M. *et al.* An HST optical to near-IR transmission spectrum of the hot Jupiter WASP-19b: detection of atmospheric water and likely absence of TiO. *Mon. Not. R. Astron. Soc.* **434**, 3252–3274 (2013).
- Mandel, K. & Agol, E. Analytic light curves for planetary transit searches. *Astrophys. J.* **580**, L171–L175 (2002).
- Pont, F., Zucker, S. & Queloz, D. The effect of red noise on planetary transit detection. *Mon. Not. R. Astron. Soc.* **373**, 231–242 (2006).
- Fortney, J. J. *et al.* A unified theory for the atmospheres of the hot and very hot Jupiters: two classes of irradiated atmospheres. *Astrophys. J.* **678**, 1419–1435 (2008).
- Burrows, A. *et al.* Photometric and spectral signatures of three-dimensional models of transiting giant exoplanets. *Astrophys. J.* **719**, 341–350 (2010).
- Seager, S. & Sasselov, D. D. Theoretical transmission spectra during extrasolar giant planet transits. *Astrophys. J.* **537**, 916–921 (2000).
- Lodders, K. Jupiter formed with more tar than ice. *Astrophys. J.* **611**, 587–597 (2004).
- Mousis, O., Lunine, J. I., Madhusudhan, N. & Johnson, T. V. Nebular water depletion as the cause of Jupiter's low oxygen abundance. *Astrophys. J.* **751**, L7 (2012).
- Wong, M. H. *et al.* Updated Galileo probe mass spectrometer measurements of carbon, oxygen, nitrogen, and sulfur on Jupiter. *Icarus* **171**, 153–170 (2004).
- Kreidberg, L. *et al.* Clouds in the atmosphere of the super-Earth GJ 1214b. *Nature* **505**, 69–72 (2014).
- Knutson, H. *et al.* *Hubble Space Telescope* near-IR transmission spectroscopy of the super-Earth HD 97658b. *Astrophys. J.* **794**, A155 (2014).
- Morley, C. V. *et al.* Neglected clouds in T and Y dwarf atmospheres. *Astrophys. J.* **756**, 172 (2012).
- Wakeford, H. R. & Sing, D. K. Transmission spectral properties of clouds for hot Jupiter exoplanets. *Astron. Astrophys.* **573**, A122 (2015).
- Showman, A. P. & Polvani, L. M. Equatorial superrotation on tidally locked exoplanets. *Astrophys. J.* **738**, 71 (2011).
- Oshagh, M. *et al.* Impact of occultations of stellar active regions on transmission spectra. *Astron. Astrophys.* **568**, A99 (2014).
- Burgasser, A. J. *et al.* Evidence of cloud disruption in the L/T dwarf transition. *Astrophys. J.* **571**, L151–L154 (2002).
- Cushing, M. C., Rayner, J. T. & Vacca, W. D. An infrared spectroscopic sequence of M, L, and T dwarfs. *Astrophys. J.* **623**, 1115–1140 (2005).

**Acknowledgements** This work is based on observations with the NASA/ESA HST, obtained at the Space Telescope Science Institute (STScI) operated by AURA, Inc. This work is also based in part on observations made with the Spitzer Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement number 336792. D.K.S., F.P. and N.N. acknowledge support from STFC consolidated grant ST/J0016/1. Support for this work was provided by NASA through grants under the HST-GO-12473 programme from the STScI. A.L.E., P.A.W. and A.V.M. acknowledge support from CNES and the French Agence Nationale de la Recherche (ANR), under programme ANR-12-BS05-0012 'Exo-Atmos'. P.A.W. and H.W. acknowledge support from the UK Science and Technology Facilities Council (STFC). G.W.H. and M.H.W. acknowledge support from NASA, NSF, Tennessee State University and the State of Tennessee through its Centers of Excellence programme.

**Author Contributions** D.K.S. led the data analysis for this project, with contributions from D.D., T.M.E., N.P.G., C.M.H., H.A.K., N.N., H.R.W., S.A., G.E.B. and P.A.W. J.J.F., A.S.B., A.P.S., A.L.E. and T.K. provided atmospheric models. G.W.H. and M.H.W. provided photometric stellar activity monitoring data and J.M.D. provided Spitzer data. D.K.S. wrote the manuscript along with J.J.F., H.R.W., T.K., N.N. and T.M.E. All authors discussed the results and commented on the draft.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.K.S. ([sing@astro.ex.ac.uk](mailto:sing@astro.ex.ac.uk)).

## METHODS

**HST observations.** The overall observational strategy is similar for each of the eight targets in the Large HST programme (GO-12473; principal investigator D.K.S.), which have been presented for WASP-19b<sup>14</sup>, HAT-P-1b<sup>12,13</sup>, WASP-12b<sup>3</sup>, WASP-31b<sup>4</sup> and WASP-6b<sup>11</sup> with the details summarized here and applied to the remaining targets HAT-P-12b, WASP-17b and WASP-39b. We observed two transits of each target with the HST STIS G430L grating, and one with the STIS G750L. The G430L and G750L data sets contain typically 43 to 53 spectra, which span either four or five spacecraft orbits and were taken with a wide 52 arcsec  $\times$  2 arcsec slit to minimize slit light losses. Both gratings have resolutions of  $R$  of  $\lambda/\Delta\lambda = 530$ –1,040 ( $\sim 2$  pixels is 5.5 Å for G430L and  $\sim 2$  pixels is 9.8 Å for G750L). The G430L grating covers the wavelength range from 2,900 Å to 5,700 Å, while the G750L grating covers 5,240 Å to 10,270 Å. The visits of HST were scheduled such that the third and/or fourth spacecraft orbits contain the transit, providing good coverage between second and third contact, as well as an out-of-transit baseline time series before and after the transit. Exposure times of 279 s were used in conjunction with a 128-pixel-wide sub-array, which reduces the readout time between exposures to 21 s, providing a 93% overall duty cycle.

The STIS data set was pipeline-reduced with the latest version of CALSTIS, and cleaned for cosmic ray detections with a customized procedure<sup>11</sup>. The G750L data set was defringed using contemporaneous fringe flats. The spectral aperture extraction was done with IRAF, using a 13-pixel-wide aperture with no background subtraction, which minimizes the out-of-transit standard deviation of the white-light curves. The extracted spectra were then Doppler-corrected to a common rest frame through cross-correlation, which helped remove sub-pixel wavelength shifts in the dispersion direction. The STIS spectra were then used to create both a white-light photometric time series and custom wavelength bands covering the spectra, integrating the appropriate wavelength flux from each exposure for different bandpasses.

Observations of HAT-P-1b and WASP-31b were also conducted in the infrared with the HST WFC3 instrument as part of GO-12473 and are detailed in refs 4 and 13. The observations use the infrared G141 grism in forward spatial scan mode over five HST orbits. Spatial scanning is done by slewing the telescope in the cross-dispersion direction during integration in a similar manner for each exposure, which increases the duty cycle and greatly increases the counts obtained per exposure. We used the ‘ima’ outputs from the CALWFC3 pipeline, which performs reference pixel subtraction, zero-read and dark current subtraction, and a nonlinearity correction. For the spectral extraction, we trimmed a wide box around each spectral image, with the spectra extracted using custom routines from the programming language IDL, similar to IRAF’s procedure from the APALL program. The aperture width was determined by minimizing the standard deviation of the fitted white-light curve. The aperture was traced around a computed centring profile, which was found to be consistent in the  $y$  axis with an error of  $<0.1$  pixels. Background subtraction was applied using a clean region of the untrimmed image. For wavelength calibration, direct images were taken in the F139M narrow-band filter at the beginning of the observations. We assumed that all pixels in the same column have the same effective wavelength, as the spatial scan varied in the  $x$ -axis direction by less than one pixel, resulting in a spectral range from 1.1  $\mu\text{m}$  to 1.7  $\mu\text{m}$ . This wavelength range was later restricted to avoid the strongly sloped edges of the grism response, which results in much lower signal-to-noise light curves.

For the comparative study, we also included the WFC3 observations for WASP-19b<sup>14</sup>, HD 209458b<sup>1</sup>, HAT-P-12b<sup>2</sup> and WASP-17b<sup>31</sup> (GO-12181; principal investigator D.D.). The WFC3 observations of WASP-12b<sup>3</sup> were also included (GO-12230; principal investigator M. R. Swain), as was HD 189733b<sup>5</sup> (GO-12881; principal investigator P. R. McCullough). The WFC3 observations of WASP-12b, WASP-17b, WASP-19b and HAT-P-12b were observed in stare mode, rather than with spatial scanning, and therefore have generally poorer overall photometric precision. See Extended Data Table 1 for a list of all observations.

**Spitzer observations.** The eight targets in the large HST survey were also all covered by Spitzer transit observations as part of an Exploration Science Programme (90092; principal investigator J.-M. Désert) obtained using the Infrared Array Camera (IRAC) instrument with the 3.6- $\mu\text{m}$  and 4.5- $\mu\text{m}$  channels in subarray mode ( $32 \times 32$  pixels). Photometry was extracted from the basic calibrated FITS data cubes, produced by the IRAC pipeline after dark subtraction, flat-fielding, linearization and flux calibration. The images contain 64 exposures taken in sequence and have per-image integration times of 1.92 s. Both channels generally show a strong ramp feature at the beginning of the time series, and we elected to trim the first  $\sim 20$  min of data to allow the detector to stabilize. We performed outlier filtering for hot (energetic) or cold (low-count values) pixels in the data by examining the time series of each pixel and subtracted the background flux from each image<sup>11</sup>.

We measured the position of the star on the detector in each image incorporating the flux-weighted centroiding method using the background subtracted pixels from each image, for a circular region with a radius of 3 pixels centred on the

approximate position of the star. We extracted photometric measurements from our data using both aperture photometry from a grid of apertures ranging from 1.5 to 3.5 pixels (in increments of 0.1) and time-variable aperture photometry. The best result was selected by measuring the flux scatter of the out-of-transit portion of the light curves for both channels after filtering the data for  $5\sigma$  outliers with a width of 20 data points.

**Transit light curve analysis.** All the transit light curves were modelled with analytical transit models<sup>15</sup>. For the white-light curves, the central transit time, orbital inclination, stellar density, planet-to-star radius contrast, stellar baseline flux and instrument systematic trends were fitted simultaneously. The period was initially fixed to a literature value, before being updated, with our final fits adopting the values obtained from an updated transit ephemeris. Both G430L transits were fitted simultaneously with a common inclination, stellar density and planet-to-star radius contrast. The results from the HST white-light curve and Spitzer fits were then used in conjunction with literature results to refine the orbital ephemeris and overall planetary system properties. To account for the effects of limb-darkening on the transit light curve, we adopted the four-parameter nonlinear limb-darkening law, calculating the coefficients with stellar models<sup>32,33</sup>.

As in our past STIS studies, we applied orbit-to-orbit flux corrections by fitting for a low-order polynomial to the photometric time series phased on the HST orbital period. The baseline flux level of each visit was free to vary in time linearly, described by two fit parameters. In addition, for the G750L we found it justified by the Bayesian Information Criteria<sup>34</sup> to also linearly fit for two further systematic trends which correlated with the  $x$  and  $y$  detector positions of the spectra, as determined from a linear spectral trace in IRAF. The orders of the fit polynomials were statistically justified based on the Bayesian Information Criteria<sup>34</sup>, and the systematic trends were fitted simultaneously with the transit parameters.

The errors on each data point were initially set to the pipeline values, which are dominated by photon noise but also includes readout noise. The best-fitting parameters were determined simultaneously with a Levenberg–Marquardt least-squares algorithm<sup>35</sup> using the unbinned data. After the initial fits, the uncertainties for each data point were rescaled based on the standard deviation of the residuals and any measured systematic errors correlated in time (‘red noise’), thus taking into account any underestimated errors calculated by the reduction pipeline in the data points. The uncertainties on the fitted parameters were calculated using the covariance matrix from the Levenberg–Marquardt algorithm, which assumes that the probability space around the best-fitting solution is well described by a multivariate Gaussian distribution and equivalent results were found when using an Markov Chain Monte Carlo analysis<sup>36</sup>. Inspection of the two-dimensional probability distributions from both methods indicated that there were no significant correlations between the planet-to-star radius contrasts and systematic trend parameters.

In an additional analysis step compared to our previous results<sup>4,11,12</sup>, we also marginalized over the systematic models<sup>37</sup> for the spectra of WASP-17b, WASP-39b, HAT-P-1b, HAT-P-12b and HD 209458b. Under this approach, we effectively averaged the results obtained from a suite of systematics models in a coherent manner. For each systematic model used to correct the data, we calculated the evidence of fit, which is then used to apply a weight to the parameter of interest ( $R_p/R_*$ ) measured using that model. In doing so, we marginalized over our uncertainty as to selecting which model is actually the ‘correct’ model. For the STIS data we included all combinations of factors up to the 4th order in both HST phase, 3rd order in detector positions  $x$  and  $y$ , 3rd order in wavelength shift, and 1st order in time. For the WFC3 data, our grid of parameterized models includes all combinations of factors up to the fourth order in both HST phase, to correct for ‘HST breathing’ effects, and up to the fourth order in wavelength shift, in addition to the visit-long linear trend. In addition, we also included exponential HST phase models, with a linear and squared planetary phase trend. For the Spitzer data, we included all combinations of the  $x$  and  $y$  positions of the stellar point spread function on the detector, including the cross-product from polynomials of  $x$  and  $y$  up to a second-order. We note that the best-fitting systematics models for HST and Spitzer are generally well constrained and the marginalized results were very similar to those based on model selection by the Bayesian Information Criteria<sup>34</sup>. For HD 209458b, lightcurve analyses and marginalization were performed using Gaussian process models<sup>38</sup>. Owing to the flexibility of Gaussian process models, a broad range of systematics behaviours can be captured without the need to provide an explicit functional form. The results of a single Gaussian process model are thus comparable to marginalizing over many simpler parametric systematics models, as was done for the other lightcurves<sup>37</sup>.

**Atmospheric models.** The synthetic spectra<sup>17,39</sup> used for this study include isothermal models as well as those with a self-consistent treatment of radiative transfer and chemical equilibrium of neutral and ionic species. Chemical mixing ratios and opacities were calculated assuming local thermochemical equilibrium accounting for condensation and thermal ionization but not photoionization<sup>40–43</sup>, for both solar metallicity and sub-solar metallicity abundances.



A simplified treatment adding in small aerosol haze particles was performed by including a Rayleigh scattering opacity (that is,  $\sigma = \sigma_0(\lambda/\lambda_0)^{-4}$ ) that had a cross-section which was  $10\times$ ,  $100\times$  and  $1,000\times$  the cross-section of molecular hydrogen gas ( $\sigma_0 = 5.31 \times 10^{-27} \text{ cm}^2$  at  $\lambda_0 = 350 \text{ nm}$ ; ref. 44). Similarly, to include the effects of a flat cloud deck we included a wavelength-independent cross-section, which was  $1\times$ ,  $10\times$  and  $100\times$  the cross-section of molecular hydrogen gas at  $350 \text{ m}$  (see Extended Data Fig. 4).

**Transmission spectral indices.** To enable a direct comparison between planets, the transmission spectra have been plotted on a common scale by dividing the measured wavelength-dependent altitude of the transmission spectra,  $z(\lambda)$ , by the planet's atmospheric scale height ( $H_{\text{eq}}$ , the vertical distance over which the gas pressure drops by a factor of  $e$ ) estimated using the equilibrium temperature. The analytical relation for the wavelength-dependent transit-measured altitude  $z(\lambda)$  of a hydrostatic atmosphere is<sup>44</sup>:

$$z(\lambda) = H \ln \left( \frac{\varepsilon P \sigma(\lambda)}{\tau} \sqrt{\frac{2\pi R_p}{kT\mu g}} \right)$$

where  $\varepsilon$  is the abundance of the absorbing or scattering species,  $P$  is the pressure at a reference altitude,  $\sigma(\lambda)$  is the wavelength-dependent cross-section,  $\tau$  is the optical thickness at the effective transit-measured radius,  $k$  is Boltzmann's constant,  $T$  is the local gas temperature,  $\mu$  is the mean mass of the atmospheric particles,  $g$  is the planetary surface gravity,  $R_p$  the planetary radius, and  $H = kT/\mu g$  is the atmospheric pressure scale height. The altitude difference measured between two wavelength regions ( $\lambda$  and  $\lambda'$ ) in a transmission spectrum is proportional to the quantity:

$$z(\lambda) - z(\lambda') = H \ln(\alpha/\alpha')$$

where  $\alpha$  is the absorption plus scattering extinction coefficient:

$$\alpha = \varepsilon \sigma(\lambda)$$

Thus, the quantity  $\Delta Z_{\lambda-\lambda'} = z(\lambda) - z(\lambda')$  is related to the ratio of the total scattering plus absorption of the atoms and molecules between the wavelength regions  $\lambda$  and  $\lambda'$ , and we use the quantity  $\Delta Z_{\lambda-\lambda'}/H_{\text{eq}} = \ln(\alpha/\alpha')$  as a metric to intercompare the atmospheric extinction for the different planets in our survey. Note that the temperature and scale height of the upper atmosphere can differ from the equilibrium value, especially at high altitudes where hot upper layers in hot Jupiters have been found<sup>45–48</sup>.

We defined indices around three main wavelength regions (see Table 1). We used a blue-optical band consisting of the G430L grating, which is sensitive between  $0.3 \mu\text{m}$  and  $0.57 \mu\text{m}$  and roughly covers the Johnson U and B photometric bandpasses. This wavelength region is almost always exclusively dominated by scattering for clear, cloudy and hazy exoplanets (see Extended Data Fig. 4). The second is a near-infrared band between  $1.22 \mu\text{m}$  and  $1.33 \mu\text{m}$ , which has overlap with the Johnson J photometric band, and is located between the strong  $\text{H}_2\text{O}$  absorption bands centred around  $1.15 \mu\text{m}$  and  $1.4 \mu\text{m}$ . This wavelength region is sensitive to the scattering continuum in hazy, cloudy and highly sub-solar models and the  $\text{H}_2\text{O}$  continuum in clear atmospheres with abundances near solar (see Extended Data Fig. 4). We also used a third wavelength region in the mid-infrared between  $3 \mu\text{m}$  and  $5 \mu\text{m}$ , which overlaps with the Johnson L and M photometric bandpasses and consists of the two Spitzer IRAC photometric channels 1 and 2. This wavelength region is highly sensitive to strong  $\text{H}_2\text{O}$ , CO and  $\text{CH}_4$  absorption bands, which are the main active molecular species expected in hot Jupiters<sup>17–19</sup>, and only sensitive to scattering in the cloudiest cases, making it an overall effective measure of the total molecular extinction (see Extended Data Fig. 4).

From the data,  $\Delta Z_{\text{UB-LM}}$  was measured taking the difference between the planet radius measured in the blue-optical HST data using the G430L grating (UB, wavelengths  $0.3\text{--}0.57 \mu\text{m}$ ) and the weighted-average value of the radii measured in Spitzer IRAC photometric channels 1 and 2 (LM, wavelengths  $3\text{--}5 \mu\text{m}$ ).  $\Delta Z_{\text{J-LM}}$  was measured similarly, although using the near-infrared WFC3 data (J, wavelengths  $1.22\text{--}1.33 \mu\text{m}$ ).

In addition, we also measured the amplitude of the near-infrared  $\text{H}_2\text{O}$  absorption band using the WFC3 spectra (see Table 1), measuring the average radii in a band containing strong  $\text{H}_2\text{O}$  absorption (between  $1.34 \mu\text{m}$  and  $1.49 \mu\text{m}$ ) compared to an adjacent band between strong  $\text{H}_2\text{O}$  features ( $1.22 \mu\text{m}$  and  $1.33 \mu\text{m}$ ). The measured  $\text{H}_2\text{O}$  amplitude for each exoplanet was then divided by the value predicted by atmospheric models<sup>17,39</sup> calculated for each planet using a planet-averaged temperature–pressure profile assuming clear atmospheres and solar abundances.

From Fig. 3, a likely inverse correlation is seen between the  $\text{H}_2\text{O}$  amplitude and the  $\Delta Z_{\text{J-LM}}/H_{\text{eq}}$  index, with the Spearman's rank correlation coefficient measured to be  $-0.76$  which has a false alarm probability of 2.8%. We note that this false alarm probability is not the probability that the water depletion scenario is correct, as that

is excluded with Fig. 3 to a much higher degree ( $5.9\sigma$  significance). A much weaker inverse correlation of  $-0.48$  is found with  $\Delta Z_{\text{UB-LM}}$  in Extended Data Fig. 2, although that has a high false alarm probability of 23%.

**Stellar activity.** As stellar activity can affect the measurement of a transmission spectrum, we photometrically monitored the activity levels of our target stars with the Cerro Tololo Inter-American Observatory (CTIO) 1.3-m telescope for the southern targets<sup>14</sup> and the Tennessee State University Celestron 14-inch (C14) Automated Imaging Telescope (AIT) located at Fairborn Observatory in Arizona for the northern targets<sup>49</sup>. All but two of our targets showed low levels of stellar activity, with observed photometric variations or upper limits which are sufficiently small that their effects on measuring the transmission spectra are minimal compared to the measurement errors<sup>3,4,11–13</sup>. The two most active stars in the survey, WASP-19A and HD 189733A, were corrected for occulted and un-occulted star spots<sup>10,14</sup>. As no contemporaneous photometric monitoring of WASP-19A is available for the July 2011 WFC3 spectra from ref. 14, we matched the spectra to the spot-corrected transit depth of  $R_p/R_* = 0.14019 \pm 0.00073$  as measured using HST WFC3 on 12 June 2014 from GO-13431 (principal investigator C. M. Huitson), which had simultaneous CTIO activity monitoring. We also normalized the differential transit depths of the WFC3 spectra<sup>5</sup> to a transit depth value consistent with ref. 10, which has a uniform treatment between the HST and Spitzer data sets of system parameters, limb-darkening and activity correction.

As effects of stellar activity could potentially mimic an optical scattering slope in a transmission spectra<sup>5,10,28,45</sup>, we searched for a relationship between the activity levels of the stars in our survey and the presence of a strong optical slope. If stellar activity were the main cause of the enhanced optical slopes, rather than scattering by hazes or clouds, then it is expected that highly active stars would have higher levels of spots and plages, and should show preferentially larger transmission spectral slopes. As an additional measure of stellar activity, we used the strength of the Ca II H and K emission lines as a stellar activity indicator ( $\log R'_{\text{HK}}$ ), as measured by Keck HIRES<sup>50,51</sup>; see Table 1. We searched for a correlation with the chromospheric activity index  $\log R'_{\text{HK}}$ , as it is correlated with the stellar photometric variability<sup>52</sup> and can be used to quantify stars with low activity levels, for which the photometric variations would be undetectable. We found no significant correlation with  $\log R'_{\text{HK}}$  activity and either the presence of haze or the strength of optical transmission spectral slope, as measured with the  $\Delta Z_{\text{J-LM}}$  index (Extended Data Fig. 3). This suggests that the effects of stellar activity are not the overall cause of the strong optical slopes seen in some of the transmission spectra.

There are also other indications that stellar activity does not have a dominant role. For one, while changing stellar activity levels should have an effect on the transmission spectra, no significant variations were seen between the three epochs of the HST STIS spectra, which has an overlapping wavelength region, for all of our targets, including active stars. In addition, the atmospheric temperature can be derived by measuring the transmission spectral slope in an atmosphere dominated by Rayleigh scattering<sup>3,4,11,45</sup>, and the temperatures found fitting a Rayleigh scattering slope for HD 189733b, HAT-P-12b and WASP-6b ( $1,340 \pm 150 \text{ K}$ ,  $1,010 \pm 80 \text{ K}$  and  $973 \pm 144 \text{ K}$ , respectively) are in good agreement with the planetary temperatures  $T_{\text{eq}}$  expected ( $1,196 \text{ K}$ ,  $958 \text{ K}$  and  $1,183 \text{ K}$ , respectively). This agreement is consistent with the atmospheric temperature, rather than stellar activity, being probed by the scattering haze. For these three stars, where HAT-P-12 has a much lower activity than the other two, the individual activity levels would have to be finely tuned for the spectral slopes to mimic the planetary temperatures.

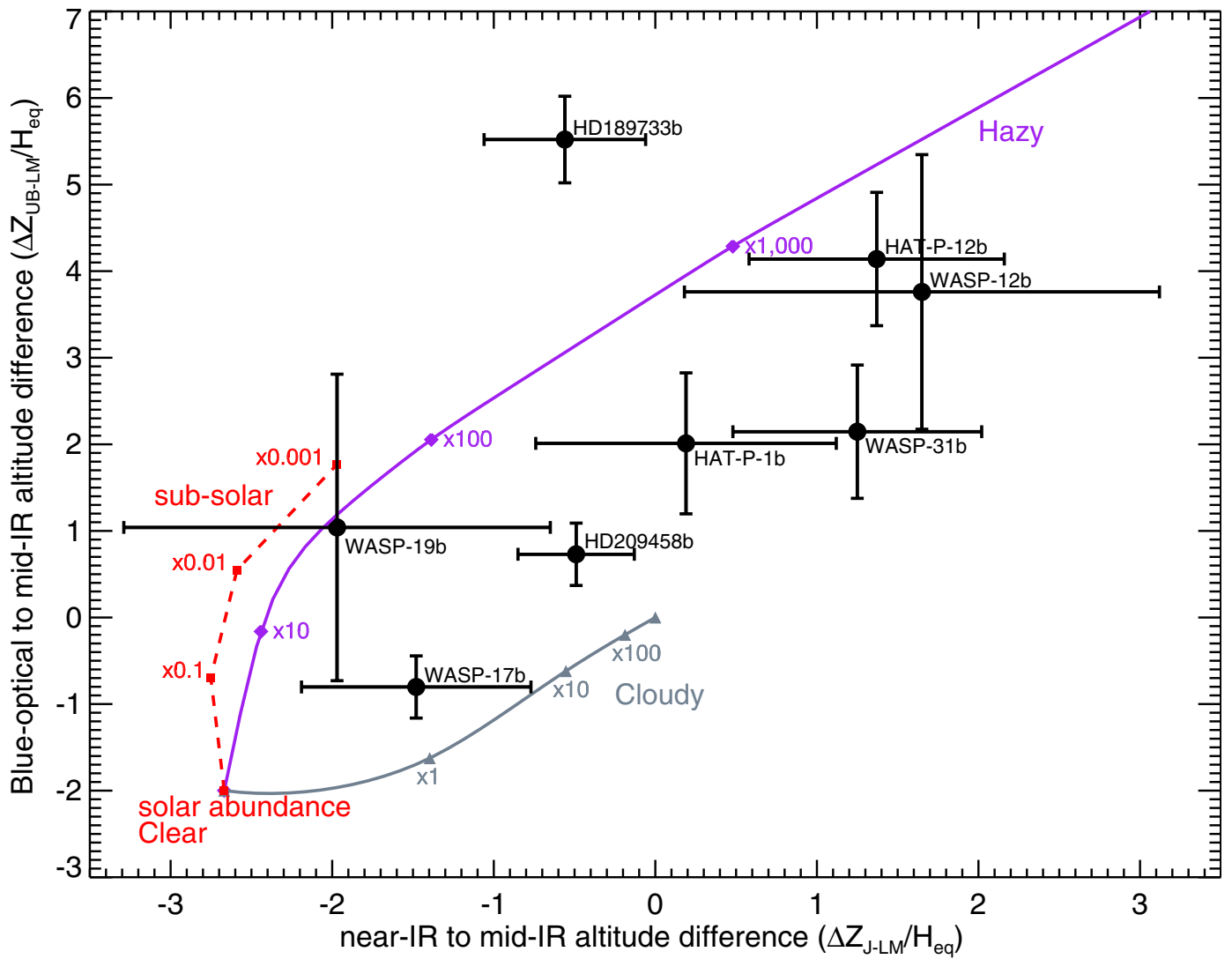
In addition to condensation chemistry<sup>53</sup>, hazes can also form through photochemical processes resulting in hydrocarbon aerosols<sup>54</sup>. This process is more effective for cooler exoplanets<sup>55</sup> and the incident stellar ultraviolet irradiation also plays an important factor in hydrocarbon formation<sup>54</sup>. Our results indicate no correlation with the presence of haze to either the atmospheric temperature or levels of ultraviolet irradiation (as traced by stellar activity indicators), which generally favours condensation chemistry over photochemical processes as the general source of the observed hazes and clouds.

**Code availability.** We have opted not to make the customized IDL codes used to produce the spectra publicly available owing to their undocumented intricacies.

31. Mandell, A. M. et al. Exoplanet transit spectroscopy using WFC3: WASP-12 b, WASP-17 b, and WASP-19 b. *Astrophys. J.* **779**, 128 (2013).
32. Sing, D. K. et al. Stellar limb-darkening coefficients for CoRoT and Kepler. *Astron. Astrophys.* **510**, A21 (2010).
33. Hayek, W., Sing, D. K., Pont, F. & Asplund, M. Limb darkening laws for two exoplanet host stars derived from 3D stellar model atmospheres. Comparison with 1D models and HST light curve observations. *Astron. Astrophys.* **539**, A102 (2012).
34. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **6**, 461–464 (1978).
35. Markwardt, C. B. Non-linear least squares fitting in IDL with MPFIT. *Astron. Soc. Pacif. Conf. Ser.* **411**, 251–254 (2009).

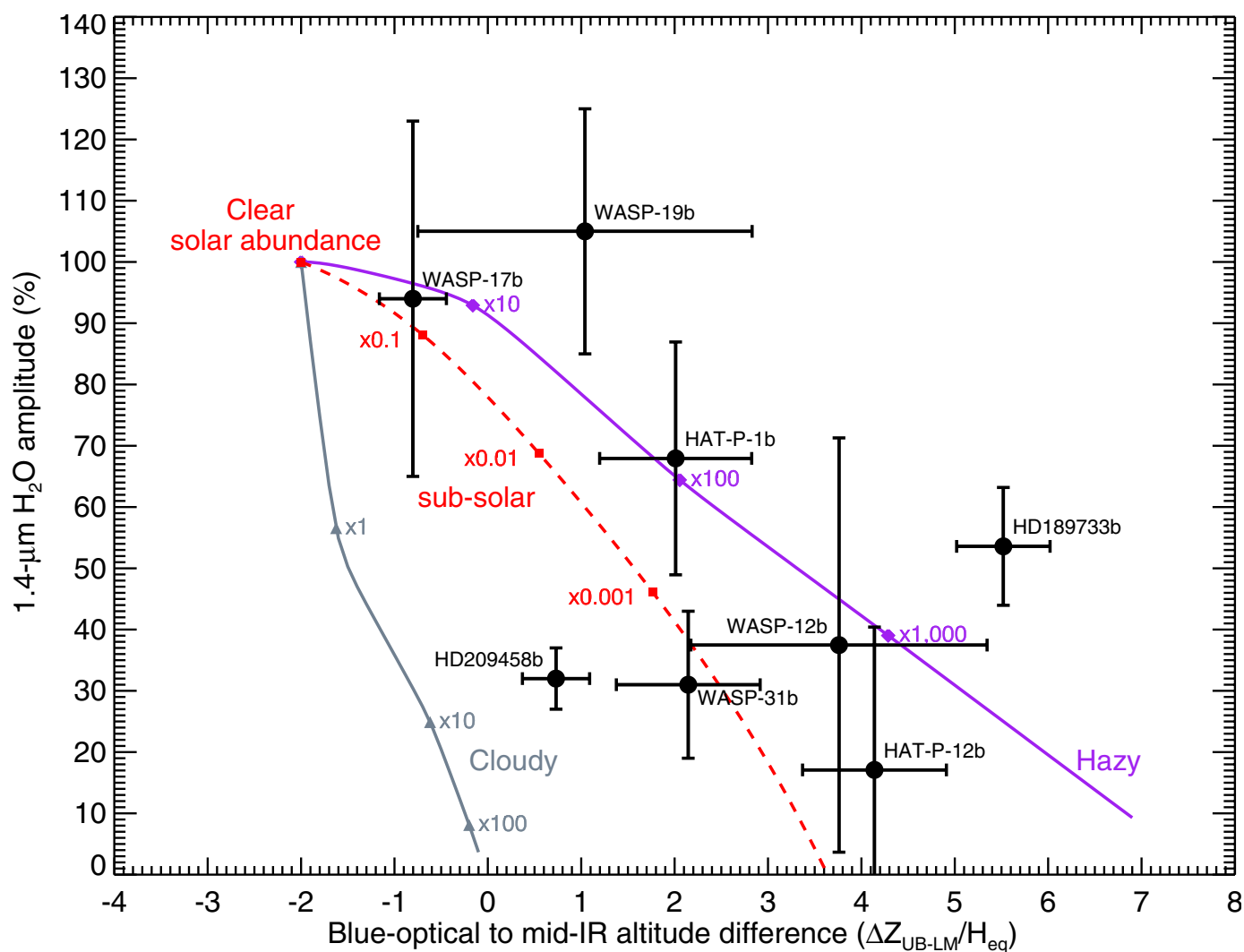


36. Eastman, J., Gaudi, B. S. & Agol, E. EXOFAST: a fast exoplanetary fitting suite in IDL. *Publ. Astron. Soc. Pacif.* **125**, 83–112 (2013).
37. Gibson, N. P. Reliable inference of exoplanet light-curve parameters using deterministic and stochastic systematics models. *Mon. Not. R. Astron. Soc.* **445**, 3401–3414 (2014).
38. Evans, T. M. *et al.* A uniform analysis of HD 209458b *Spitzer*/IRAC lightcurves with Gaussian process models. *Mon. Not. R. Astron. Soc.* **451**, 680–694 (2015).
39. Fortney, J. J. *et al.* Transmission spectra of three-dimensional hot Jupiter model atmospheres. *Astrophys. J.* **709**, 1396–1406 (2010).
40. Lodders, K. Alkali element chemistry in cool dwarf atmospheres. *Astrophys. J.* **519**, 793–801 (1999).
41. Lodders, K. & Fegley, B. Atmospheric chemistry in giant planets, brown dwarfs, and low-mass dwarf stars. I. Carbon, nitrogen, and oxygen. *Icarus* **155**, 393–424 (2002).
42. Freedman, R. S., Marley, M. S. & Lodders, K. Line and mean opacities for ultracool dwarfs and extrasolar planets. *Astrophys. J.* **174** (Supp.), 504–513 (2008).
43. Visscher, C., Lodders, K. & Fegley, B. Jr. Atmospheric chemistry in giant planets, brown dwarfs, and low-mass dwarf stars. III. Iron, magnesium, and silicon. *Astrophys. J.* **716**, 1060–1075 (2010).
44. Lecavelier des Etangs, A. *et al.* Rayleigh scattering in the transit spectrum of HD 189733b. *Astron. Astrophys.* **481**, L83–L86 (2008).
45. Sing, D. K. *et al.* *Hubble Space Telescope* transmission spectroscopy of the exoplanet HD 189733b: high-altitude atmospheric haze in the optical and near-ultraviolet with STIS. *Mon. Not. R. Astron. Soc.* **416**, 1443–1455 (2011).
46. Huitson, C. M. *et al.* Temperature–pressure profile of the hot Jupiter HD 189733b from *HST* sodium observations: detection of upper atmospheric heating. *Mon. Not. R. Astron. Soc.* **422**, 2477–2488 (2012).
47. Vidal-Madjar, A. *et al.* The upper atmosphere of the exoplanet HD 209458 b revealed by the sodium D lines. Temperature–pressure profile, ionization layer, and thermosphere. *Astron. Astrophys.* **527**, A110 (2011).
48. Wyttenbach, A., Ehrenreich, D., Lovis, C., Udry, S. & Pepe, F. Spectrally resolved detection of sodium in the atmosphere of HD 189733b with the HARPS spectrograph. *Astron. Astrophys.* **577**, A62 (2015).
49. Henry, G. W. Techniques for automated high-precision photometry of Sun-like stars. *Publ. Astron. Soc. Pacif.* **111**, 845–860 (1999).
50. Isaacson, H. & Fischer, D. Chromospheric activity and jitter measurements for 2630 stars on the California Planet Search. *Astrophys. J.* **725**, 875–885 (2010).
51. Knutson, H. A., Howard, A. W. & Isaacson, H. A correlation between stellar activity and hot Jupiter emission spectra. *Astrophys. J.* **720**, 1569–1576 (2010).
52. Lockwood, G. W. *et al.* Patterns of photometric and chromospheric variation among sun-like stars: a 20 year perspective. *Astrophys. J.* **171**, 260–303 (2007).
53. Helling, Ch., Woitke, P. & Thi, W.-F. Dust in brown dwarf and extra-solar planets I. Chemical composition and spectral appearance of quasi-static cloud layers. *Astron. Astrophys.* **485**, 547–560 (2008).
54. Liang, M.-C. *et al.* On the insignificance of photochemical hydrocarbon aerosols in the atmospheres of close-in extrasolar giant planets. *Astrophys. J.* **605**, L61–L64 (2004).
55. Moses, J. I. *et al.* Disequilibrium carbon, oxygen, and nitrogen chemistry in the atmospheres of HD 189733b and HD 209458b. *Astrophys. J.* **737**, 15 (2011).



**Extended Data Figure 1** |  $\Delta Z_{J-LM}$  versus  $\Delta Z_{UB-LM}$ . Black points show the altitude difference between the near-infrared (near-IR) and the mid-infrared (mid-IR) spectral features ( $\Delta Z_{J-LM}$ ) versus the difference between the blue-optical and mid-infrared ( $\Delta Z_{UB-LM}$ ), see Table 1). Error bars represent the  $1\sigma$  measurement uncertainties. Purple and grey

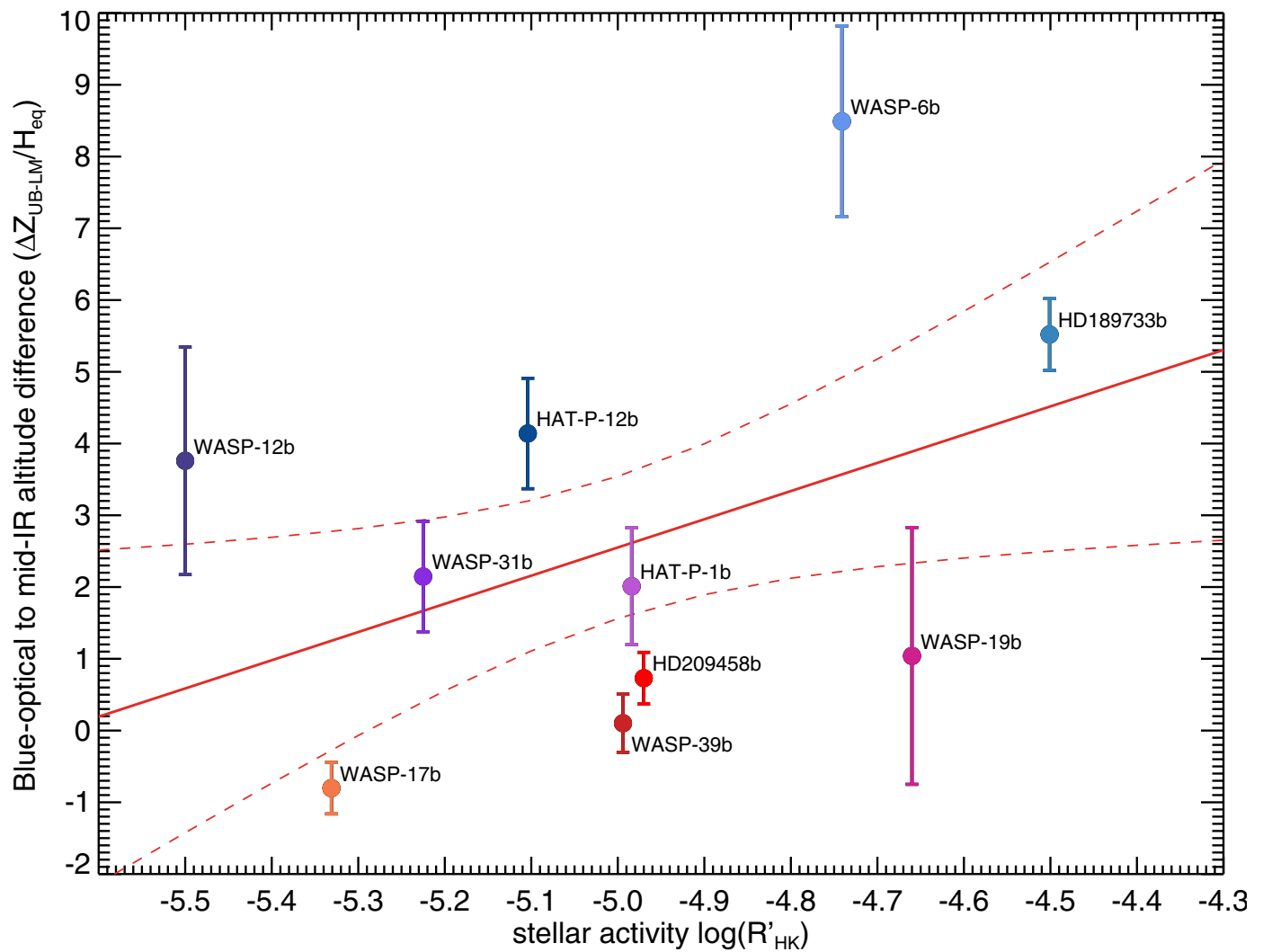
lines show model trends for hazy and cloud atmospheres, respectively, with increasing Rayleigh scattering haze and grey cloud deck opacity corresponding to  $10\times$ ,  $100\times$  and  $1,000\times$  solar. We also show clear-atmosphere models with sub-solar abundances of  $0.1\times$ ,  $0.01\times$  and  $0.001\times$  solar (red line).



**Extended Data Figure 2 |  $\Delta Z_{UB-LM}$  index versus H<sub>2</sub>O amplitude.**

Black points show the altitude difference between the blue-optical and mid-infrared (mid-IR) spectral features ( $\Delta Z_{UB-LM}$ ) versus the amplitude of the 1.4-μm H<sub>2</sub>O absorption spectral feature (see Table 1). Error bars represent the 1σ measurement uncertainties. Purple and grey lines show

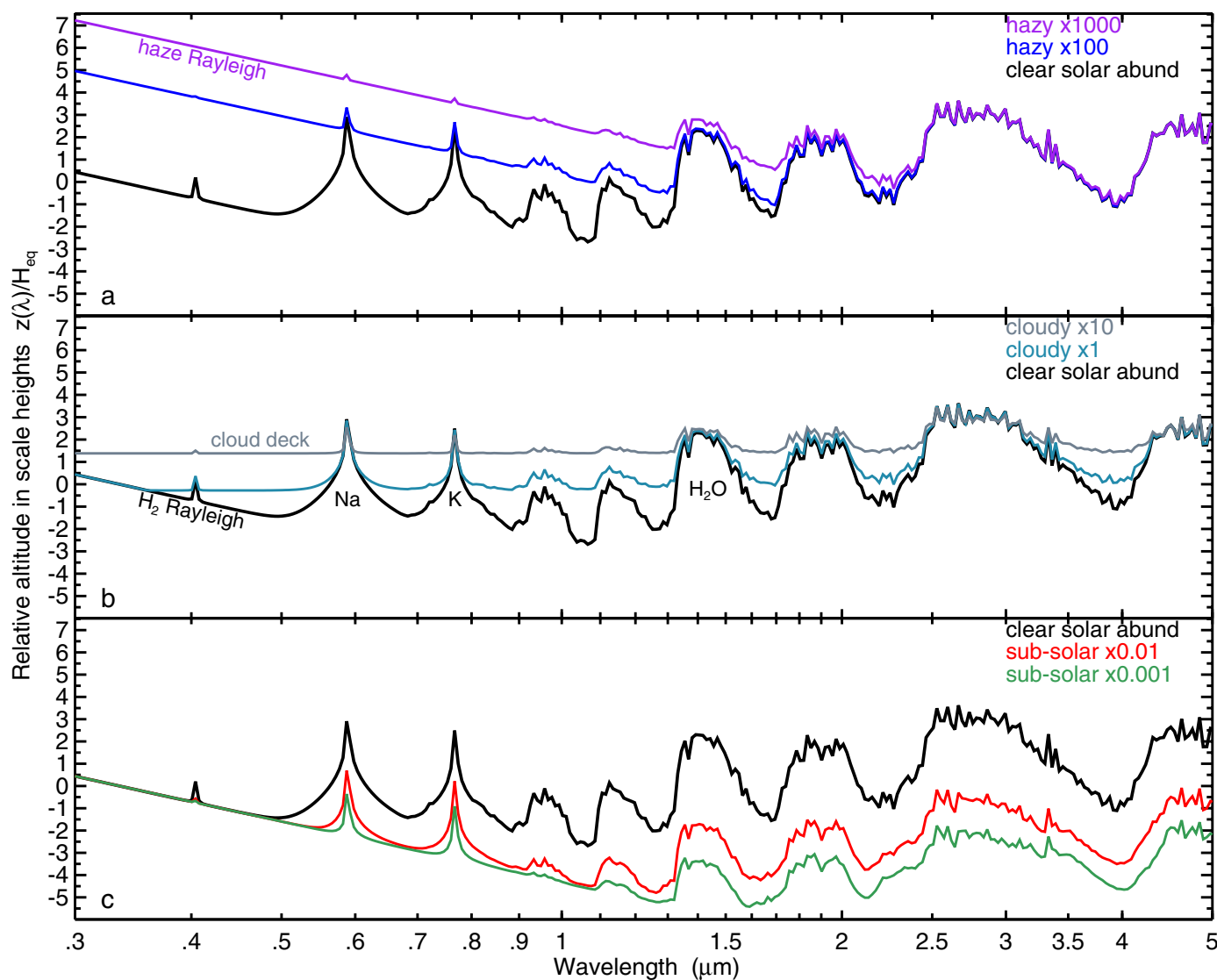
model trends for hazy and cloud atmospheres, respectively, with increasing Rayleigh scattering haze and grey cloud deck opacity corresponding to 10×, 100× and 1,000× solar. We also show clear-atmosphere models with sub-solar abundances of 0.1×, 0.01× and 0.001× solar (red line).



**Extended Data Figure 3 | Stellar activity ( $\log R'_{HK}$ ) versus  $\Delta Z_{UB-LM}$  index.** Exoplanets with strong haze signatures have prominent optical slopes with  $\Delta Z_{UB-LM}$  values above 3, while clear atmospheres have

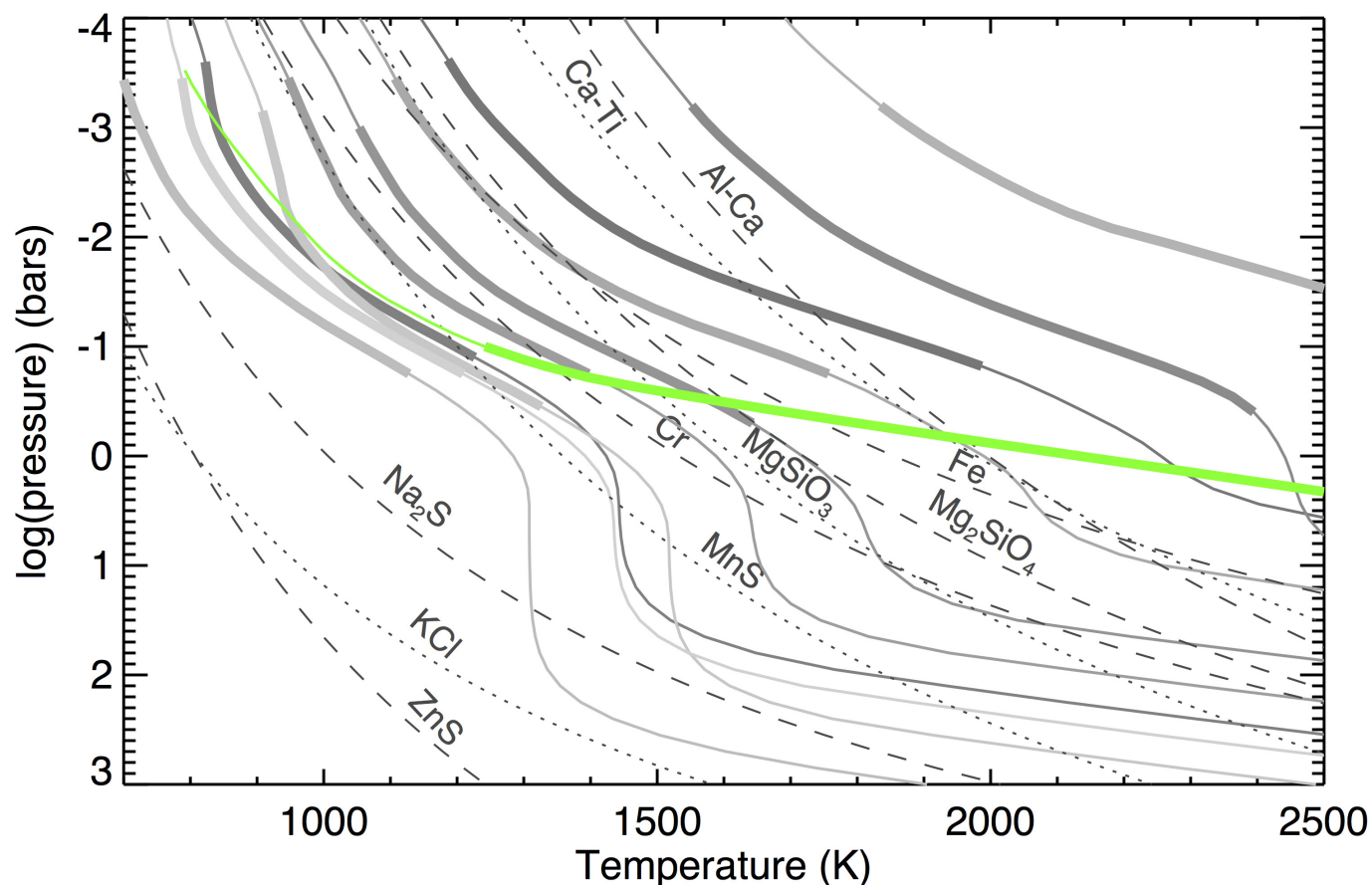
$\Delta Z_{UB-LM}$  indices near zero. The datapoint colours correspond to those in Fig. 1. The red solid line shows the linear regression between the two indices, with  $1\sigma$  uncertainties (red dashed lines).





**Extended Data Figure 4 | Theoretical model transmission spectra.** Model spectra<sup>17,39</sup> assume a 1,200-K hot Jupiter with a surface gravity of  $25 \text{ m s}^{-2}$ . Spectra in each panel are compared to a clear, solar-metallicity atmosphere (black line). **a**, Purple spectra have an added Rayleigh

scattering haze corresponding to metallicities of 100 $\times$  and 1,000 $\times$  solar. **b**, Blue and grey spectra have an added grey cloud deck corresponding to 1 $\times$  and 10 $\times$  solar. **c**, Red and green spectra show clear atmospheres with sub-solar abundances of 0.01 $\times$  and 0.001 $\times$  solar.



**Extended Data Figure 5 | Brown dwarf and hot Jupiter pressure–temperature profiles and condensation curves.** Similar to Fig. 2, but alongside the ten hot-Jupiter pressure–temperature profiles we plot the profile of an 1,800-K brown dwarf (green line). The thicker portions of the lines indicate the pressures probed in transmission for the hot Jupiters

(plotted in greyscale) and the visible photosphere for the brown dwarf (0.1–10 bar). While a shift in the pressure–temperature profile of a hot Jupiter to hotter and cooler temperatures could dramatically change which condensates may be found in the visible atmosphere, the same would not be true for much shallower brown dwarf pressure–temperature profiles.

Extended Data Table 1 | Summary of observations

Planet	V mag.	Optical <i>HST</i> STIS, ACS 0.3 – 1 $\mu$ m grating, Date (UT)	near-IR <i>HST</i> WFC3 1.1 – 1.7 $\mu$ m grism, Date (UT)	mid-IR <i>Spitzer</i> IRAC 3 – 5.2 $\mu$ m band, Date (UT)
HAT-P-1b	9.87	G430L, 2012/05/26 G750L, 2012/05/30 G430L, 2012/09/19	G141, 2012/07/05	3.6, 2013/09/11 4.5, 2013/09/20
WASP-6b	11.91	G430L, 2012/06/10 G430L, 2012/06/16 G750L, 2012/07/23	—	4.5, 2013/01/14 3.6, 2013/01/21
WASP-12b	11.57	G430L, 2012/03/14 G430L, 2012/03/27 G750L, 2012/09/09	G141, 2011/04/12	3.6, 2010/11/17 4.5, 2010/12/11
WASP-31b	11.98	G430L, 2012/06/13 G430L, 2012/06/26 G750L, 2012/07/10	G141, 2012/05/13	3.6, 2013/09/13 4.5, 2013/03/19
HAT-P-12b	12.84	G430L, 2012/04/11 G430L, 2012/04/30 G750L, 2013/02/04	G141, 2011/05/29	3.6, 2013/03/08 4.5, 2013/03/11
WASP-19b	12.59	G430L, 2012/04/30 G430L, 2012/05/04 G750L, 2012/05/09	G141, 2011/07/01 G141, 2014/06/12	3.6, 2011/08/03 4.5, 2011/08/13
WASP-39b	12.09	G430L, 2013/02/08 G430L, 2013/02/12 G750L, 2013/03/17	—	3.6, 2013/04/18 4.5, 2013/10/10
WASP-17b	11.83	G430L, 2012/06/08 G430L, 2013/03/15 G750L, 2013/03/19	G141, 2011/07/08	4.5, 2013/05/10 3.6, 2013/05/14
HD 209458b	7.63	G430L, 2003/05/03 G750L, 2003/05/31 G430L, 2003/06/25 G750L, 2003/07/05 G750M, 2000/04/25 G750M, 2000/04/28 G750M, 2000/05/05 G750M, 2000/05/12	G141, 2012/09/25	3.6, 2007/12/31 3.6, 2008/07/19 3.6, 2011/01/14 3.6, 2014/01/19 4.5, 2008/07/22 4.5, 2010/01/19
HD 189733b	7.648	G800L, 2006/05/22 G800L, 2006/05/26 G800L, 2006/07/14 G430L, 2009/11/20 G430L, 2010/05/18 G750M, 2009/10/30 G750M, 2009/11/13 G750M, 2010/09/30 G750M, 2010/11/29	G141, 2013/06/05	3.6, 2007/11/25 3.6, 2006/10/30 3.6, 2010/12/29 4.5, 2007/11/22 4.5, 2009/12/23

Transit observations using the Hubble and Spitzer Space Telescopes. Dates are given in universal time (UT) listed along with the instruments and wavelength ranges used.

# Evidence for a new phase of dense hydrogen above 325 gigapascals

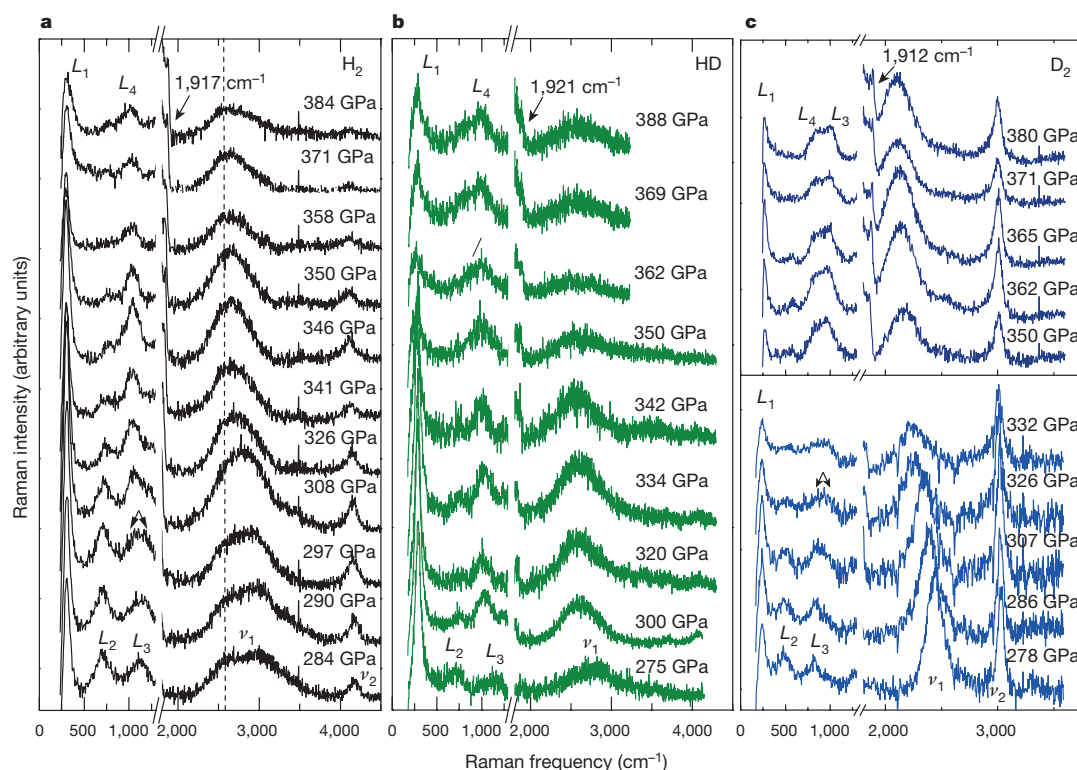
Philip Dalladay-Simpson<sup>1</sup>, Ross T. Howie<sup>1†</sup> & Eugene Gregoryanz<sup>1,2</sup>

Almost 80 years ago it was predicted that, under sufficient compression, the H–H bond in molecular hydrogen (H<sub>2</sub>) would break, forming a new, atomic, metallic, solid state of hydrogen<sup>1</sup>. Reaching this predicted state experimentally has been one of the principal goals in high-pressure research for the past 30 years. Here, using *in situ* high-pressure Raman spectroscopy, we present evidence that at pressures greater than 325 gigapascals at 300 kelvin, H<sub>2</sub> and hydrogen deuteride (HD) transform to a new phase—phase V. This new phase of hydrogen is characterized by substantial weakening of the vibrational Raman activity, a change in pressure dependence of the fundamental vibrational frequency and partial loss of the low-frequency excitations. We map out the domain in pressure–temperature space of the suggested phase V in H<sub>2</sub> and HD up to 388 gigapascals at 300 kelvin, and up to 465 kelvin at 350 gigapascals; we do not observe phase V in deuterium (D<sub>2</sub>). However, we show that the transformation to phase IV' in D<sub>2</sub> occurs above 310 gigapascals and 300 kelvin. These values represent the largest known isotropic shift in pressure, and hence the largest possible pressure difference between the H<sub>2</sub> and D<sub>2</sub> phases, which implies that the appearance of phase V of D<sub>2</sub> must occur at a pressure of above 380 gigapascals. These experimental data provide a glimpse of the physical properties of

dense hydrogen above 325 gigapascals and constrain the pressure and temperature conditions at which the new phase exists. We speculate that phase V may be the precursor to the non-molecular (atomic and metallic) state of hydrogen that was predicted 80 years ago.

The exchange interaction, a purely quantum mechanical effect, forms one of the strongest bonds in chemistry—the H–H bond. Owing to this bond, hydrogen exists in molecular form, with atoms separated by approximately 0.74 Å and a bond dissociation energy of approximately 4.52 eV (refs 2, 3) at ambient conditions. The first experiments to break this bond<sup>4</sup> demonstrated that extreme conditions are needed to do so; for example, the H<sub>2</sub> molecule dissociates only to a minor extent at high temperatures (at 3,000 K, the degree of dissociation is around 10%)<sup>5</sup>. Another mechanism to break the hydrogen bond—pressure—was subsequently proposed<sup>1</sup>; it was theorized that above 250,000 atm (25 GPa), the hydrogen molecules would dissociate, forming solid, atomic, metallic hydrogen, an entirely new state of the first and simplest element.

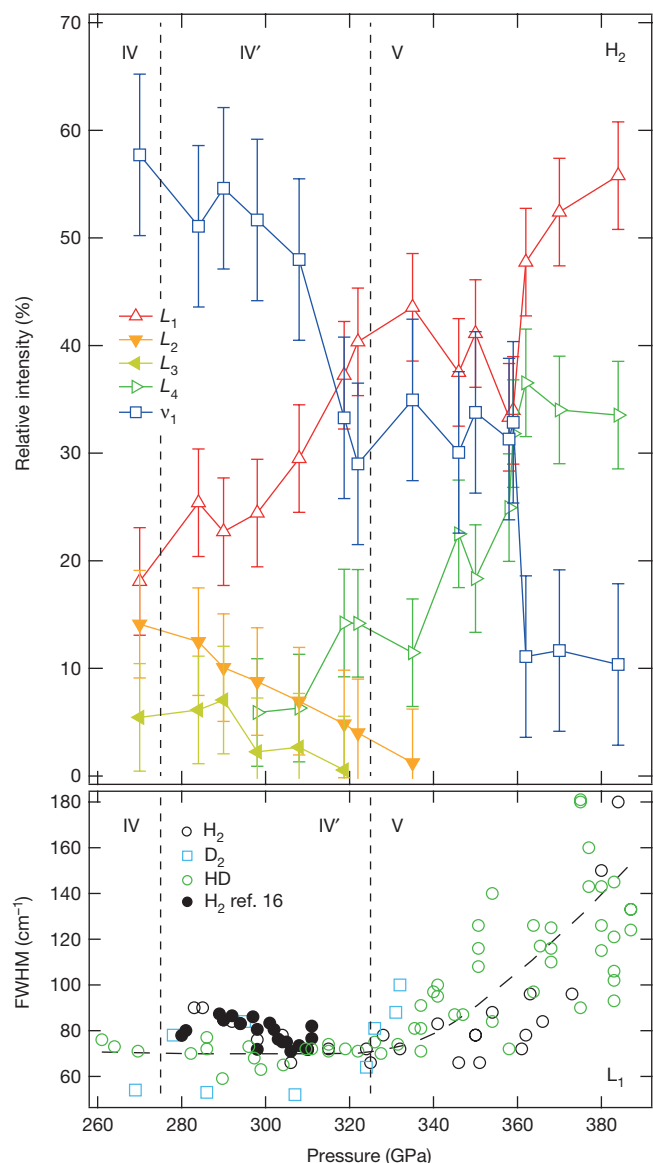
The proposed high-pressure route to an atomic metallic state has proved to be one of the great experimental challenges in high-pressure physics. Despite the technological advances in high-pressure physics, this theoretical prediction has yet to be experimentally confirmed, even



**Figure 1 | Representative Raman spectra of three hydrogen isotopes.** a–c, Spectra of hydrogen are shown in black (a), hydrogen deuteride in green (b) and deuterium in blue (c). The spectra at different pressures (as labelled) are plotted as a waterfall; the offset is for clarity. The low-frequency modes  $L_{1-4}$  and the vibrational modes  $\nu_{1,2}$  are labelled. The second-order diamond band spanning 2,300–2,600  $\text{cm}^{-1}$  is visible on some spectra; its central position is marked by the dashed vertical line in a. The diamond edge for each isotope, which was used to determine the pressure, is labelled on the highest-pressure spectra (1,917  $\text{cm}^{-1}$  for H<sub>2</sub>, 1,921  $\text{cm}^{-1}$  for HD and 1,912  $\text{cm}^{-1}$  for D<sub>2</sub>). HD was formed from a mixture of H<sub>2</sub> (75%) and D<sub>2</sub> (25%); see Methods. The spectra were collected using a 647.1-nm excitation wavelength. The arrows in a and c represent the splitting of the  $L_3$  mode and the onset of IV' (for clarity see Extended Data Fig. 4).

<sup>1</sup>School of Physics and Centre for Science at Extreme Conditions, University of Edinburgh, Edinburgh EH9 3JZ, UK. <sup>2</sup>Key Laboratory of Materials Physics, Institute of Solid State Physics, Chinese Academy of Sciences, Hefei 230031, China. †Present address: Center for High Pressure Science & Technology Advanced Research, Shanghai 201203, China.





**Figure 2 | Relative intensities of the vibrational, low-frequency modes and the full-width at half-maximum of the  $L_1$  mode as a function of pressure.** **a**, Relative intensities of the vibrational ( $\nu_1$ ) and four low-frequency modes ( $L_{1-4}$ ) of hydrogen represented as a percentage of the total Raman activity of the sample; error bars reflect the accuracy of the measurement (see Methods and Extended Data Fig. 7). The low-frequency modes  $L_2$  and  $L_3$  disappear at around 325 GPa. **b**, The full-width at half-maximum (FWHM) of the low-frequency mode  $L_1$  of H<sub>2</sub>, D<sub>2</sub> and HD as function of pressure; the dashed curve is a guide to the eye. The dashed vertical lines in **a** and **b** indicate the transformations to phases IV' and V.

at pressures (and high temperatures) an order of magnitude higher than that originally proposed<sup>6–14</sup>. Recently, a new solid phase of dense hydrogen—phase IV—was experimentally discovered<sup>15,16</sup> at 300 K and above 230 GPa. This new phase IV exhibits a change in the gradient of the fundamental vibrational-mode frequency  $\nu_1$  with respect to pressure  $P$  at a constant temperature  $T = 300$  K,  $(d\nu_1/dP)_T$ , which leads to extremely low values of  $\nu_1$  above 230 GPa; for example,  $\nu_1 \approx 2,750 \text{ cm}^{-1}$  at 315 GPa (ref. 16). This value is indicative of a much weaker bond, compared to ambient conditions, and is consistent with the bond length of approximately  $0.82 \text{ \AA}$  (ref. 17). It was observed that phase IV could be viewed as a mixed molecular and atomic state and that the complete dissociation of the hydrogen molecule is feasible at even higher compressions<sup>18</sup>.

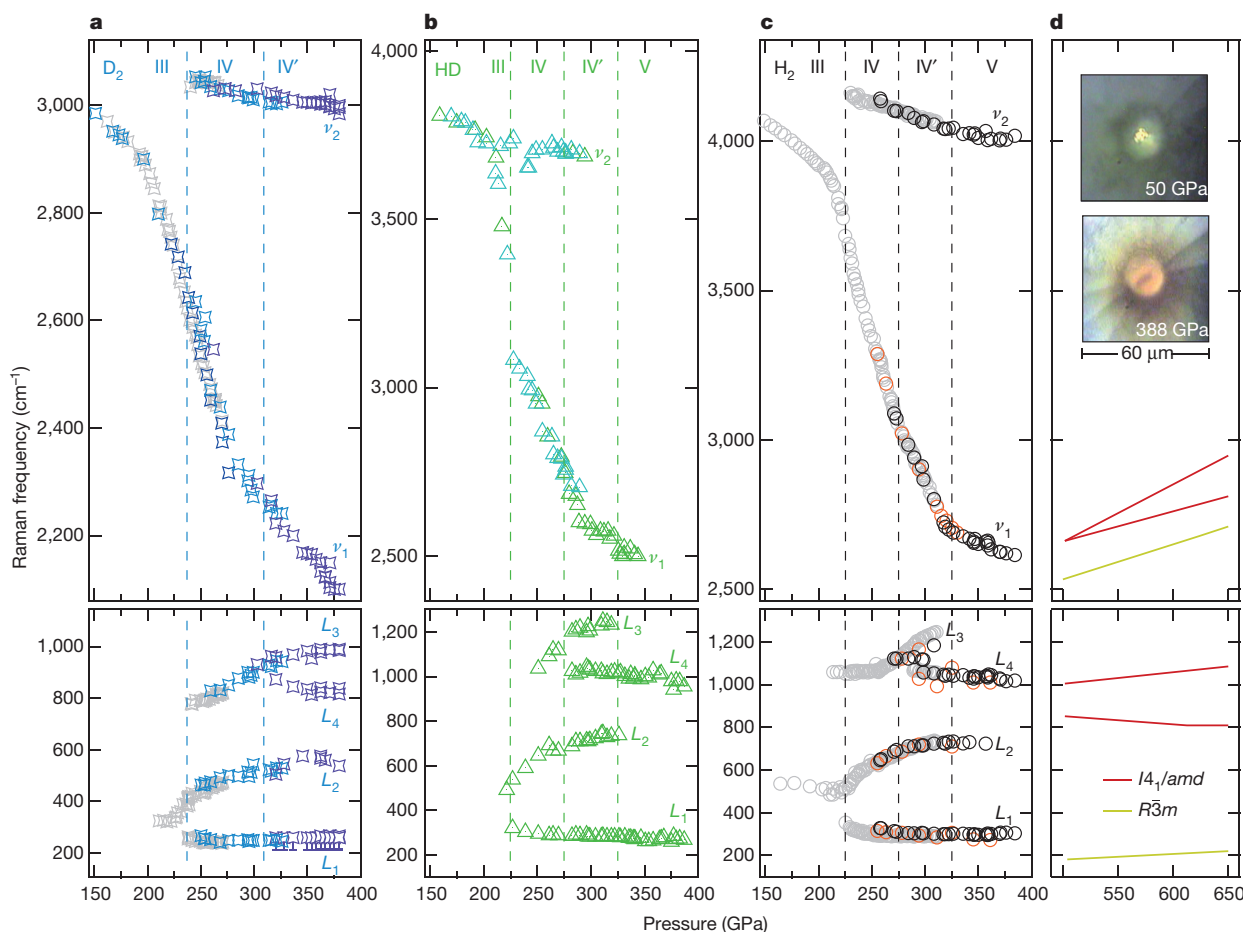
To investigate the states of hydrogen above 320 GPa, we conducted very high pressure studies on H<sub>2</sub>, HD (hydrogen deuteride) and D<sub>2</sub>,

reaching pressures of  $384 \pm 15$  GPa,  $388 \pm 15$  GPa and  $380 \pm 15$  GPa, respectively. These pressures, despite being conservative estimates (see Methods and Extended Data Figs 1 and 2), are still among the highest pressures reported so far in a diamond anvil cell, and the highest pressures hydrogen has so far been subjected to in static experiments. On the basis of the substantial decrease in intensity of vibrational Raman bands, the change of slope of the vibrational-mode frequency with pressure, and changes in position, width and intensity of the low-frequency ( $<1,300 \text{ cm}^{-1}$ ) modes, we tentatively infer a transition to a new structural configuration—phase V—of H<sub>2</sub> and HD above 325 GPa, while observing phase IV' of D<sub>2</sub> above 310 GPa. We present experimental information on the physical properties of the dense hydrogen just below 400 GPa and provide some constraints on the  $P$ – $T$  space of phase V. On the basis of the optical changes observed through the phase transition, we speculate that the proposed phase V might be the onset of a non-molecular state of hydrogen.

Figure 1 shows the representative Raman spectra of three isotopes of hydrogen compressed at 300 K. (For the full description of the relevant experimental details, see Methods and refs 18 and 19; further information about the intensities of the modes and frequencies of HD as a function of H/D concentration and pressure is provided in ref. 20.) Above 220 GPa, all hydrogen isotopes enter phase IV, which is characterized by sharp, well-defined, low-frequency modes (Fig. 1, marked for clarity as  $L_1$ ,  $L_2$  and  $L_3$  in all figures; see also ref. 20) and the presence of a second vibrational fundamental mode  $\nu_2$ . The appearance of the Raman spectra of HD (at similar pressures) is essentially identical to those of H<sub>2</sub> or D<sub>2</sub> (refs 16, 18); see Extended Data Fig. 3. When pressures above 275 GPa are reached (for H<sub>2</sub> and HD), we observe a change in the gradient of the frequency with respect to pressure of the  $L_3$  mode, and its branching to produce a new  $L_4$  mode. These changes mark the appearance of phase IV', described previously<sup>18</sup>. It was suggested that phase IV' could structurally resemble phase IV, on the basis of close similarities between the Raman spectra<sup>18</sup>. Above 320 GPa we observe gradual, but profound, modification in the Raman spectra, indicative of the phase transformation to a new phase, phase V (H<sub>2</sub> and HD only). The pressure needed to enter phase IV' in deuterium is 35 GPa higher, as evidenced by the splitting of  $L_3$  into  $L_3$  and  $L_4$  at 310 GPa (Fig. 1, Extended Data Fig. 4).

In hydrogen, after branching to produce the  $L_4$  mode, the  $L_3$  mode slowly redistributes its intensity into the  $L_4$  mode (Figs 1, 2 and Extended Data Fig. 4). When the suggested phase V is reached, the  $L_3$  mode completely disappears and the intensity of  $L_4$  becomes comparable to that of the  $L_1$  mode (Figs 1 and 2). Meanwhile, the  $L_1$  mode undergoes a marked change itself; Fig. 2b shows the full-width at half-maximum (FWHM) of  $L_1$  as function of pressure. At the same pressure as when the vibrational Raman modes start to become weaker and the  $L_2$  and  $L_3$  modes disappear ( $>325$  GPa; see Figs 1 and 2), the FWHM of the  $L_1$  mode starts to increase rapidly. Between 330 GPa and 388 GPa the width of the  $L_1$  mode increases more than twofold, reaching  $180 \text{ cm}^{-1}$  by 388 GPa (Fig. 2b). Even though the  $L_1$  mode is very broad at the highest pressures, it remains the dominant feature of the spectra of all isotopes (Fig. 1). We also observe some small but detectable softening of the  $L_1$  frequency with pressure (Fig. 3).

Up to 325 GPa, the total Raman intensity of all modes stays roughly the same (Fig. 2) for all three isotopes, in agreement with previous studies<sup>16</sup> of pure H<sub>2</sub> up to 315 GPa. However, when pressures above 325 GPa are reached, the low-frequency modes  $L_2$  and  $L_3$  disappear and the intensities of both vibrational excitations of H<sub>2</sub> and HD start to decrease rapidly. In the case of hydrogen, the  $\nu_2$  modes become almost indistinguishable from the background above 358 GPa, whereas the  $\nu_1$  mode becomes broad and weak, overlapping with the second-order diamond band (Fig. 1a); the positions of the hydrogen and deuterium vibrational modes are clearly visible in all spectra (Fig. 1). The second-order diamond mode spanning the approximate range  $2,300$ – $2,600 \text{ cm}^{-1}$  overlaps in frequency with the  $\nu_1$  mode of all isotopes,



**Figure 3 | Frequencies of the vibrational and low-energy modes of the isotopes as functions of pressure.** **a–c**, The data for the different isotopes are shown as open stars ( $D_2$ ; **a**), triangles with centre dots (HD; **b**) and open circles ( $H_2$ ; **c**), with different colours representing different experimental runs. Data from previous studies<sup>16,18</sup> are shown as grey symbols. The vertical dashed lines denote the phase transitions in the

which makes the estimation of its intensity difficult. In the case of HD, it became impossible to distinguish between the second-order diamond mode and the  $\nu_1$  mode above 350 GPa (Fig. 1b). The notable decrease of the vibrational-mode intensity means that the spectra of the suggested phase V looks highly unusual, particularly when compared with those of phase IV, in which the vibrational mode dominates (see >270–320-GPa spectra of all isotopes in Fig. 1). As well as the pronounced drop of the intensities of the vibrational modes, we observe a change in the slope of the  $\nu_1$  frequency with pressure ( $d\nu_1/dP$ )<sub>T</sub> at around 325 GPa for hydrogen and hydrogen deuteride (Fig. 3). The  $\nu_1$  mode softens rapidly with pressure in phase IV (average gradient of  $-12\text{ cm}^{-1}\text{ GPa}^{-1}$ ; ref. 16) and changes to a rate of about  $-7\text{ cm}^{-1}\text{ GPa}^{-1}$  (refs 16, 18) in phase IV'. When hydrogen is compressed to more than 325 GPa, the softening of the  $\nu_1$  mode essentially stops and ( $d\nu_1/dP$ )<sub>T</sub> becomes almost independent of pressure (equal to  $-1.37\text{ cm}^{-1}\text{ GPa}^{-1}$ ), resembling that of the  $\nu_2$  mode ( $-1.01\text{ cm}^{-1}\text{ GPa}^{-1}$ ; see Fig. 3). The change of slope and the sudden increase of the FWHM of the  $L_1$  mode happen at the same pressure, suggesting that the nature of the bonding is noticeably modified by the transition between phase IV(IV') and V'. In a recent Raman optical study<sup>21</sup>, a small change in the slope of the vibrational mode of the  $H_2$  vibrational mode at 300 GPa was observed, from which three structural phase transitions within 50 GPa (275–325 GPa) were inferred. However, our data do not seem to support these findings (Extended Data Fig. 5).

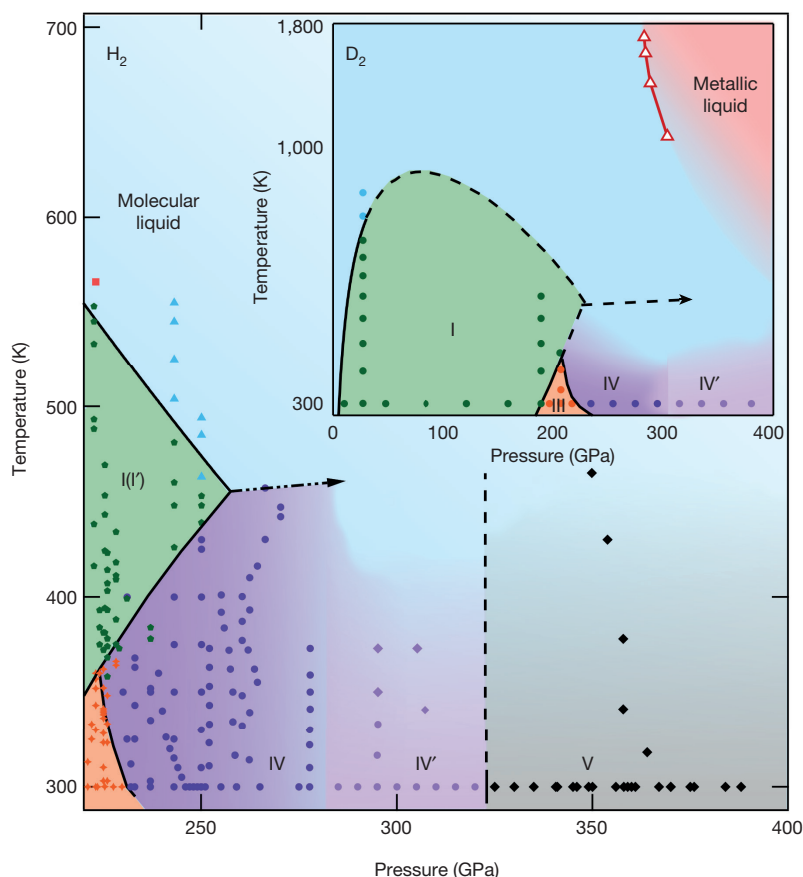
The pressure at which phase IV of deuterium appears is about 10 GPa higher than that of hydrogen<sup>18</sup>, whereas the transition from phases IV

corresponding isotope. **d**, The theoretically calculated frequencies of hydrogen for the metallic and non-molecular (atomic) structures of  $I4_1/amd$  (red) and  $R\bar{3}m$  (yellow) from ref. 22. The insets are photos of the HD sample at 50 GPa and at 388 GPa, as labelled, taken in transmitted and reflected light.

to IV' is shifted by 35–40 GPa. We observe similar qualitative changes in the slope of the deuterium vibrational mode at 310 GPa upon entrance into phase IV', but the slope remains relatively steep, resulting in the extremely low vibrational frequency of approximately  $2,100\text{ cm}^{-1}$  at 380 GPa. The large pressure difference between phase IV' of hydrogen and that of deuterium suggests that phase V of deuterium will appear at pressures above 380 GPa.

We investigated the  $P$ – $T$  space where phase IV(IV') and the proposed phase V exist by conducting heating experiments. If hydrogen is heated at 250 GPa, then the phase IV  $\leftrightarrow$  I transformation happens at 430 K, and at 450 K phase I presumably melts (see Fig. 4 and the figures in ref. 14). In some runs, phase IV(IV') was heated at pressures above the I–IV–liquid<sup>14</sup> triple point—for example, at 262 GPa to approximately 450 K and at 270 GPa and 290–310 GPa to approximately 375 K—but no transformations to phase V were observed (Fig. 4). Finally, we heated phase V at 350 GPa and did not observe any transformation up to 465 K (Extended Data Fig. 6). These points in  $P$ – $T$  space indicate that phase V is separated from the lower-pressure phase IV(IV') by a phase line that is probably close to vertical (Fig. 4).

The decrease in the vibrational-mode intensities could indicate the loss of sample, particularly in the case of hydrogen, but the observations described above rule this out and instead indicate a possible phase transition. These observations include: the evolution of the low-frequency modes (that is, broadening and frequency change) with pressure up to 390 GPa, and only a modest drop in the intensity of the  $L_1$  mode; the noticeable change in the slope of the vibrational-mode



**Figure 4 | Proposed phase diagram of hydrogen up to 400 GPa.** The coloured, filled symbols and solid phase lines below 300 GPa in the main figure are from ref. 14, and show phases I(I'), III and IV(IV'). The solid black diamonds (phase V) are from this study, the vertical, grey dashed line indicates the transition from phase IV' to phase V and the dashed-dotted arrow is the proposed continuation of the melting curve. The inset shows a sketch of the phase diagram of D<sub>2</sub>. The coloured, filled symbols and solid lines were obtained by us in another, unreported study. The dashed lines are the proposed melting curves of deuterium, which have not been measured experimentally, but are assumed to follow the same trend as those of hydrogen. The red open triangles are from ref. 26 and separate the metallic and semi-conducting liquids.

frequency with pressure at 325 GPa at 300 K; and the lack of sample loss or detection of transformation upon heating at pressures above 320 GPa. In heating experiments, rapid sample loss is observed in the liquid state, which results in the complete disappearance of all hydrogen Raman activity and the resulting spectra resemble those of the gasket (Extended Data Fig. 6).

It is tempting, although highly speculative at this time, to interpret phase V as the onset of the predicted<sup>1</sup> non-molecular and metallic state of hydrogen. *Ab initio* random-structure searches that included zero-point motion estimate that hydrogen should dissociate into atomic and metallic states at around 500 GPa (ref. 22) and 380 GPa (ref. 23), respectively. The possible lowest-energy structural candidates include the tetragonal  $I4_1/amd$  and trigonal  $R\bar{3}m$  symmetries<sup>22,23</sup>. Both structures have inter-atomic Raman phonons with frequencies of about  $2,500\text{ cm}^{-1}$  at 500 GPa, which is close to the frequency of the vibrational mode  $\nu_1$  of hydrogen that we observe at 380 GPa (see Fig. 3 and supplementary information in ref. 22). In calculations, these phonons are present up to 4.2 TPa, slowly increasing in energy with increasing pressure<sup>22</sup> as the distance between the atoms decreases. The presence of the extremely weak  $\nu_2$  mode at 384 GPa (Fig. 1a) indicates that the purely atomic state was not reached in our experiments and that slightly higher pressures are required to completely dissociate hydrogen. It is plausible that the molecular dissociation commences at pressures above 350 GPa, resulting in the alterations of the Raman spectrum as described here. If the suggested phase V is indeed the beginning of the complete molecular dissociation of partially molecular phase IV', then it could explain all the optical observations presented here, such as the band gap decreasing with pressure (1.8 eV at 315 GPa in phase IV', ref. 16). Furthermore, the possible appearance of conducting electrons due to dissociation could explain the very dark appearance of the sample as seen in transmitted and reflected light in the visible region (Fig. 3d, inset), and the overall decrease of the Raman intensities. The relatively simple overall Raman spectrum observed experimentally matches those predicted theoretically rather well (Fig. 3d). The  $I4_1/amd$

symmetry does not predict a very prominent  $L_1$  mode, whereas the  $R\bar{3}m$  symmetry, which is more energetically favourable at even higher pressures, does not predict the  $L_4$  mode (see Fig. 3b, c), both of which are observed experimentally. However, these discrepancies could be accounted for by the 100-GPa pressure range between theory and experiment. A minimum in  $(d\nu_1/dP)_T$  would indicate the evolution from the intramolecular vibrational mode to an interatomic phonon. This change could require a 100-GPa pressure range to complete and would result in hardening of the phonons at pressures above 500 GPa.

The data from this and a previous melting study<sup>14</sup> provide further insight into the current phase diagram of hydrogen (Fig. 4). It appears that there could be another triple point between the proposed phase V, phase IV(IV'), and a liquid state (not shown) at above 275 GPa and 450 K. If phase V is indeed a precursor to a fully non-molecular, and presumably metallic, solid state, then a question arises about the existence and location of the phase line separating the molecular (insulating) and non-molecular (metallic) liquids and solids. A non-molecular liquid could be expected to exist in the same pressure range as phase V, but at higher temperatures. In fact, theoretical studies have suggested a phase transition from a molecular liquid to an atomic liquid in hydrogen<sup>24,25</sup>. The data presented in refs 24 and 25 suggest the existence of the highly conducting atomic liquid state at pressures as low as about 150 GPa and above 2,000 K (ref. 25). However, shock-wave experiments<sup>26</sup> indicate the existence of the metallic liquid deuterium at higher pressures of 350 GPa, with the corresponding phase line being almost vertical (Fig. 4, inset). These experimental results seem to be in a very good agreement with our current study. Extrapolation of the data from ref. 26 to lower temperatures would imply yet another triple point between the melting curve and the two liquid phases. The presence of two dissimilar liquids would suggest the presence of two solid phases below them, with properties mimicking those of the liquids, for example, non-molecular (insulating) versus atomic (metallic). Experimental confirmation of the location of the phase line(s) and triple points would be very

important for the complete description of phase V, even higher-pressure solid phases and the possible molecular–atomic transition. An understanding of the connection between the proposed metallization and phase V is also required. Such additional data could provide invaluable information about the fundamental physics and chemistry that governs the behaviour of the simplest element at high densities.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 16 June; accepted 14 October 2015.**

- Wigner, E. & Huntington, H. B. On the possibility of a metallic modification of hydrogen. *J. Chem. Phys.* **3**, 764–770 (1935).
- van Kranendonk, J. *Solid Hydrogen* (Plenum Press, 1983).
- Housecroft, C. E. & Sharpe, A. G. *Inorganic Chemistry* (Prentice Hall, 2007).
- Langmuir, I. & Mackay, G. M. J. The dissociation of hydrogen into atoms. Part I. Experimental. *J. Am. Chem. Soc.* **36**, 1708–1722 (1914).
- Langmuir, I. The dissociation of hydrogen into atoms. [Part II.] Calculation of the degree of dissociation and the heat of formation. *J. Am. Chem. Soc.* **37**, 417–458 (1915).
- Goncharov, A. F., Gregoryanz, E., Hemley, R. J. & Mao, H.-k. Spectroscopic studies of the vibrational and electronic properties of solid hydrogen to 285 GPa. *Proc. Natl Acad. Sci. USA* **98**, 14234–14237 (2001).
- Loubeyre, P., Occelli, F. & LeToullec, R. Optical studies of solid hydrogen to 320 GPa and evidence for black hydrogen. *Nature* **416**, 613–617 (2002).
- Gregoryanz, E., Goncharov, A., Matsuishi, K., Mao, H.-k. & Hemley, R. Raman spectroscopy of hot dense hydrogen. *Phys. Rev. Lett.* **90**, 175701 (2003).
- Eremets, M. I. & Trojan, I. A. Evidence of maximum in the melting curve of hydrogen at megabar pressures. *JETP Lett.* **89**, 174–179 (2009).
- Subramanian, N., Goncharov, A. F., Struzhkin, V. V., Somayazulu, M. & Hemley, R. J. Bonding changes in hot fluid hydrogen at megabar pressures. *Proc. Natl Acad. Sci. USA* **108**, 6014–6019 (2011).
- Akahama, Y. *et al.* Evidence from x-ray diffraction of orientational ordering in phase III of solid hydrogen at pressures up to 183 GPa. *Phys. Rev. B* **82**, 060101 (2010).
- Akahama, Y., Kawamura, H., Hirao, N., Ohishi, Y. & Takemura, K. Raman scattering and x-ray diffraction experiments for phase III of solid hydrogen. *J. Phys. Conf. Ser.* **215**, 012056 (2010).
- Zha, C.-S., Liu, Z. & Hemley, R. J. Synchrotron infrared measurements of dense hydrogen to 360 GPa. *Phys. Rev. Lett.* **108**, 146402 (2012).
- Howie, R. T., Dalladay-Simpson, P. & Gregoryanz, E. Raman spectroscopy of hot hydrogen above 200 GPa. *Nature Mater.* **14**, 495–499 (2015).
- Eremets, M. I. & Troyan, I. A. Conductive dense hydrogen. *Nature Mater.* **10**, 927–931 (2011).
- Howie, R. T., Guillaume, C. L., Scheler, T., Goncharov, A. F. & Gregoryanz, E. Mixed molecular and atomic phase of dense hydrogen. *Phys. Rev. Lett.* **108**, 125501 (2012).
- Pickard, C. J. & Needs, R. J. Structure of phase III of solid hydrogen. *Nature Phys.* **3**, 473–476 (2007).
- Howie, R. T., Scheler, T., Guillaume, C. L. & Gregoryanz, E. Proton tunneling in phase IV of hydrogen and deuterium. *Phys. Rev. B* **86**, 214104 (2012).
- Howie, R. T., Gregoryanz, E. & Goncharov, A. F. Hydrogen (deuterium) vibron frequency as a pressure comparison gauge at multi-Mbar pressures. *J. Appl. Phys.* **114**, 073505 (2013).
- Howie, R. T., Magdau, I. B., Goncharov, A. F., Ackland, G. J. & Gregoryanz, E. Phonon localization by mass disorder in dense hydrogen-deuterium binary alloy. *Phys. Rev. Lett.* **113**, 175501 (2014).
- Zha, C.-s., Cohen, R. E., Mao, H.-k. & Hemley, R. J. Raman measurements of phase transitions in dense solid hydrogen and deuterium to 325 GPa. *Proc. Natl Acad. Sci. USA* **111**, 4793–4797 (2014).
- McMahon, J. M. & Ceperley, D. M. Ground-state structures of atomic metallic hydrogen. *Phys. Rev. Lett.* **106**, 165302 (2011).
- Azadi, S., Monserrat, B., Foulkes, W. M. C. & Needs, R. J. Dissociation of high-pressure solid molecular hydrogen: a quantum Monte Carlo and anharmonic vibrational study. *Phys. Rev. Lett.* **112**, 165501 (2014).
- Tamblyn, I. & Bonev, S. A. Structure and phase boundaries of compressed liquid hydrogen. *Phys. Rev. Lett.* **104**, 065702 (2010).
- Morales, M., Pierleoni, C., Schwegler, E. & Ceperley, D. M. Evidence for a first-order liquid-liquid transition in high-pressure hydrogen from ab initio simulations. *Proc. Natl Acad. Sci. USA* **107**, 12799–12803 (2010).
- Knudson, M. D. *et al.* Direct observation of an abrupt insulator-to-metal transition in dense liquid deuterium. *Science* **348**, 1455–1460 (2015).

**Acknowledgements** We are grateful to M. Frost for assistance during experiments. This work is supported by a Leadership Fellowship from the UK Engineering and Physical Sciences Research Council (EPSRC), reference number EP/J003999/1. P.D-S. acknowledges studentship funding from EPSRC grant number EP/G03673X/1.

**Author Contributions** R.T.H. and P.D-S. carried out the experiments, analysed the data and wrote the paper. E.G. conceived and designed the project, carried out the experiments, analysed the data and wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.G. (e.gregoryanz@ed.ac.uk).



## METHODS

**Sample loadings.** The experimental runs used mostly the same techniques and method described in refs 16, 18, 19 and references therein. For this study we conducted a total of 14 independent experiments up to pressures of  $388 \pm 15$  GPa. In some of the runs we heated the sample, at different pressures, up to temperatures of 465 K. Pressure was generated in long, high-temperature, piston-cylinder diamond anvil cells of our own design equipped with diamonds with culet dimensions ranging from  $30 \mu\text{m}$  to  $15 \mu\text{m}$ . The rhenium foils with thicknesses of  $200\text{--}250 \mu\text{m}$  were used as the gasket material to form the sample chamber. The hydrogen gas was clamped at  $0.175\text{--}0.200$  GPa at 300 K and then further compressed to above 150 GPa, usually within 2–3 h after clamping. The HD was produced by mixing the pure isotopes in gas phases (usually  $<10$  MPa) at 300 K. The partial pressures were used to calculate the composition, which, for the experiment on HD described here, was 75% and 25% for hydrogen and deuterium, respectively (see also ref. 20).

**Optical measurements.** We used 514.15-nm and 647.1-nm excitation wavelengths to collect the spectra. Owing to the quantum efficiency of the visible CCD (charged coupled device) used, the high-energy modes—for example, hydrogen vibrational excitation at above  $3,500 \text{ cm}^{-1}$ —are much weaker than the low-energy lattice modes if probed using a 647.1-nm wavelength. However, in most of the cases, when 514.15-nm excitation is used, the pressure-induced fluorescence from the stressed diamonds obscures the Raman signal, which leaves 647.1-nm excitation as the only available source, as in Fig. 1.

**Pressure and temperature measurements.** For pressure measurements, the stressed-diamond-edge frequency was used and, where applicable, cross-referenced with the frequency of the vibrational modes<sup>19</sup> from previous experiments to maintain self-consistency. An example of how the frequency of the stressed diamond edge was determined, and the dependence of the vibrational frequency of hydrogen versus the frequency of the stressed diamond edge is given in Extended Data Fig. 1a. The first-order diamond Raman band becomes elongated in frequency space, composed of two sharp, well-defined peaks: one corresponding to the stressed culet and the other to the unstressed regions of the diamond. The frequency from the stressed culet was determined by the frequency ( $\omega$ ) at which  $dI/d\omega$  was minimized (where  $I$  is the intensity of the spectrum), a technique proposed in refs 11, 27, and 28.

The calibration data presented in refs 27 and 28 were primarily used here for determining pressure. These two curves agree up to about 200 GPa, but gradually diverge at higher pressures (Extended Data Fig. 2b). For example, at the highest pressure reached, we observed a diamond-edge frequency of  $1,936 \text{ cm}^{-1}$  (see Fig. 1), which corresponds to pressures of approximately 388 GPa and 403 GPa on the scales proposed by Akahama & Kawamura in 2004<sup>27</sup> and 2006<sup>28</sup>, respectively.

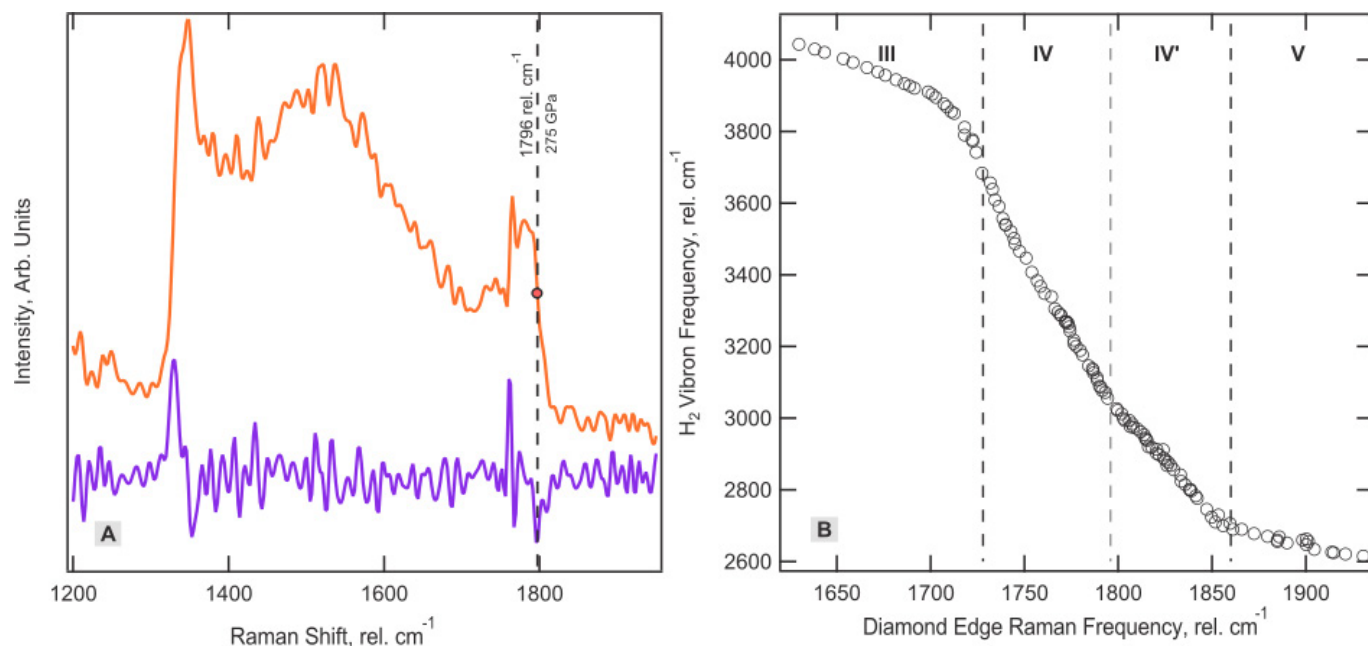
With their latest calibration in 2010<sup>29</sup>, this frequency corresponds to a substantially higher pressure of 449 GPa (Extended Data Fig. 2b). However, the effect of pressures above about 300 GPa on soft samples has yet to be determined; the latest calibration<sup>29</sup> up to 410 GPa needs to be independently verified, particularly for softer samples. To be consistent with previous results, we decided to use the most conservative scale<sup>27</sup>, as was used in our previous studies<sup>16,18,19</sup>. This scale provides a smooth continuation of the frequencies of the low-energy and vibrational modes versus pressure observed by us in all experiments.

We therefore stress that the characteristics that provide evidence for the phase V transition are independent of the choice of the previously discussed calibrations, not a direct consequence. Extended Data Figure 2a demonstrates that the discontinuous change in  $d\nu_1/dP$  for pure  $\text{H}_2$  is present when using any of the stressed-diamond-edge pressure calibrations, and remains just as prominent when using the less-conservative and more-contemporary pressure scales<sup>11,27</sup>.

For heating, we used two custom-built resistive heaters placed around the diamonds and the body of the cell. Temperature was determined using one or two thermocouples, attached to one of the diamonds and/or the gasket.

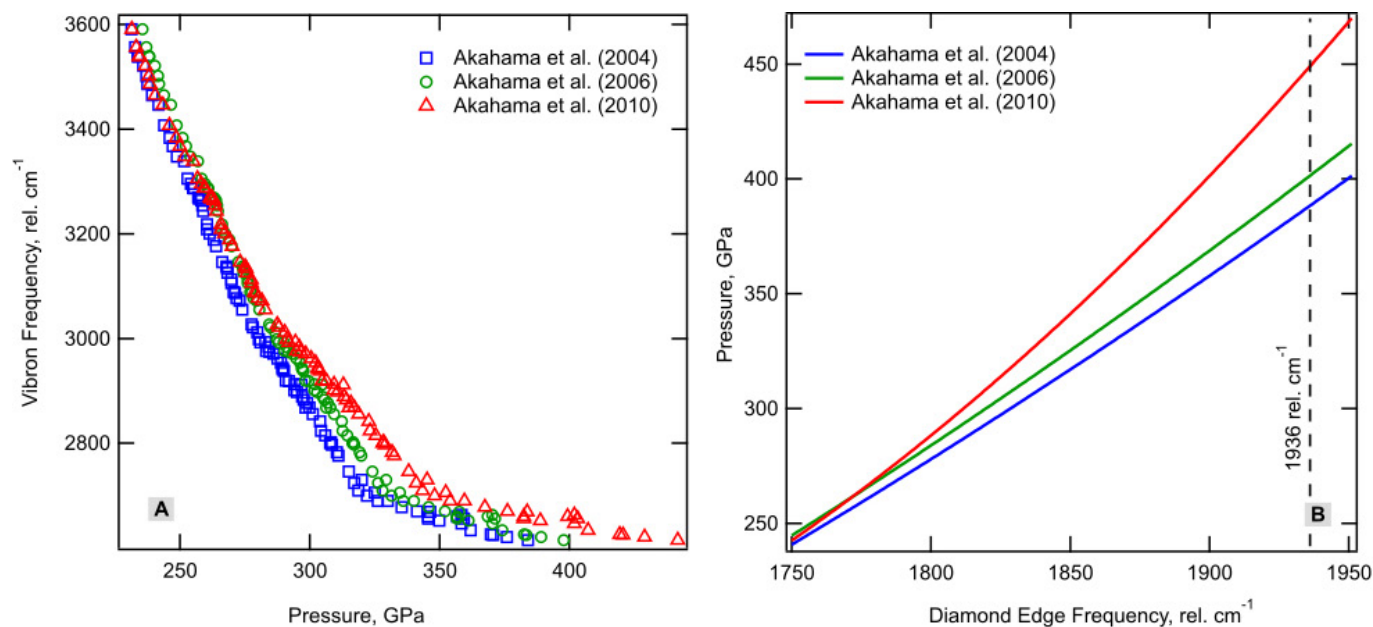
**Calculating relative integrated intensities.** Calculating the relative Raman intensities in the diamond anvil cell is a difficult task, especially when these intensities are of similar magnitude to the relatively low signal-to-noise ratios. Therefore, the data in Fig. 2 are from the spectrum with the highest signal-to-noise ratio in each run. First, the background caused by the pressure-induced fluorescence of the diamond anvils is subtracted. The residual data are then fitted with Voigt profiles, which produces values for the integrated intensities of each excitation. These values are then summed, and the percentage of total Raman activity is calculated; an example is provided in Extended Data Fig. 7. Owing to the extremely small samples, the second-order Raman band also becomes comparable in magnitude to the excitations from the sample (Extended Data Fig. 7, inset). Consequently, at higher pressures for which the  $\nu_1$  excitations overlap with the second-order Raman diamond band, extra care has to be taken. Here, the evolution of the spectra with pressure (which is determined using fits from a previous pressure step as an initial guess) as well as the relationship between the intensity of the first- and second-order diamond bands are used to accurately determine the integrated intensity of  $\nu_1$ .

27. Akahama, Y. & Kawamura, H. High-pressure Raman spectroscopy of diamond anvils to 250 GPa: method for pressure determination in the multimegabar pressure range. *J. Appl. Phys.* **96**, 3748–3751 (2004).
28. Akahama, Y. & Kawamura, H. Pressure calibration of diamond anvil Raman gauge to 310 GPa. *J. Appl. Phys.* **100**, 043516 (2006).
29. Akahama, Y. & Kawamura, H. Pressure calibration of diamond anvil Raman gauge to 410 GPa. *J. Phys. Conf. Ser.* **215**, 012195 (2010).



**Extended Data Figure 1 | Calculating pressure.** **a**, A typical example of a spectrum from the first-order Raman band of diamond when probing the sample (orange). The frequency edge is given by the vertical dashed line at 1,796  $\text{rel. cm}^{-1}$ , which corresponds to a pressure of 275 GPa (ref. 27).

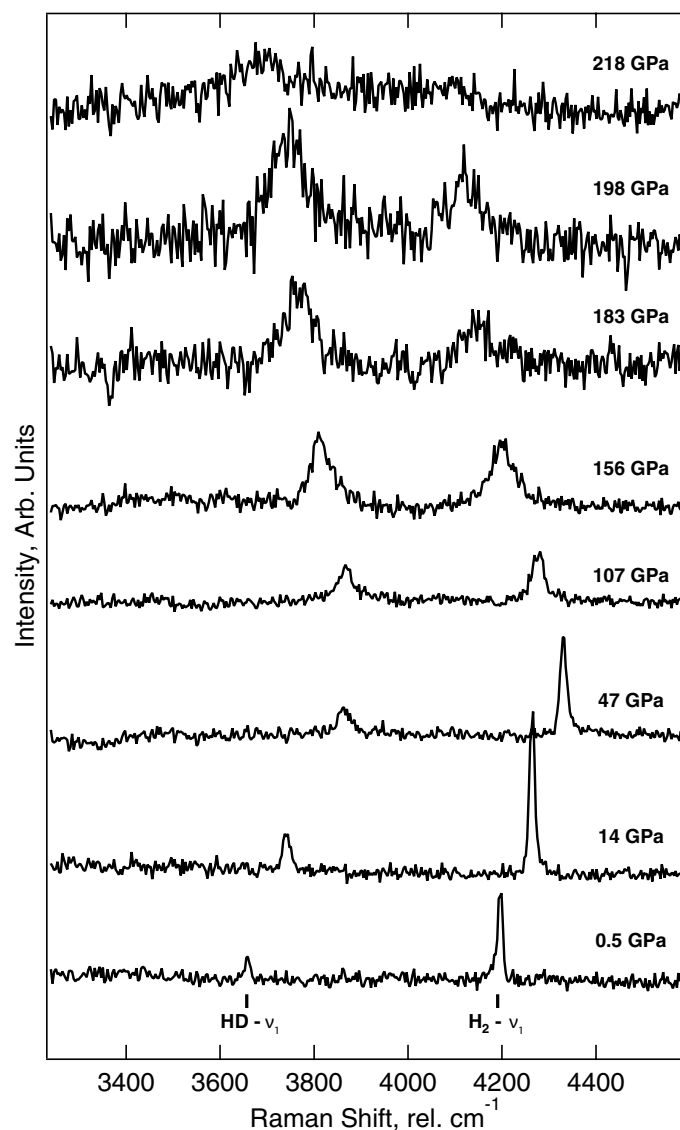
This stressed edge is defined as the frequency that minimizes  $dI/d\omega$  (purple). **b**,  $\text{H}_2$  vibrational-mode (vibron) frequency ( $\nu_1$ ) plotted as a function of the stressed-diamond-edge frequency.



### Extended Data Figure 2 | Comparison of pressure calibrations.

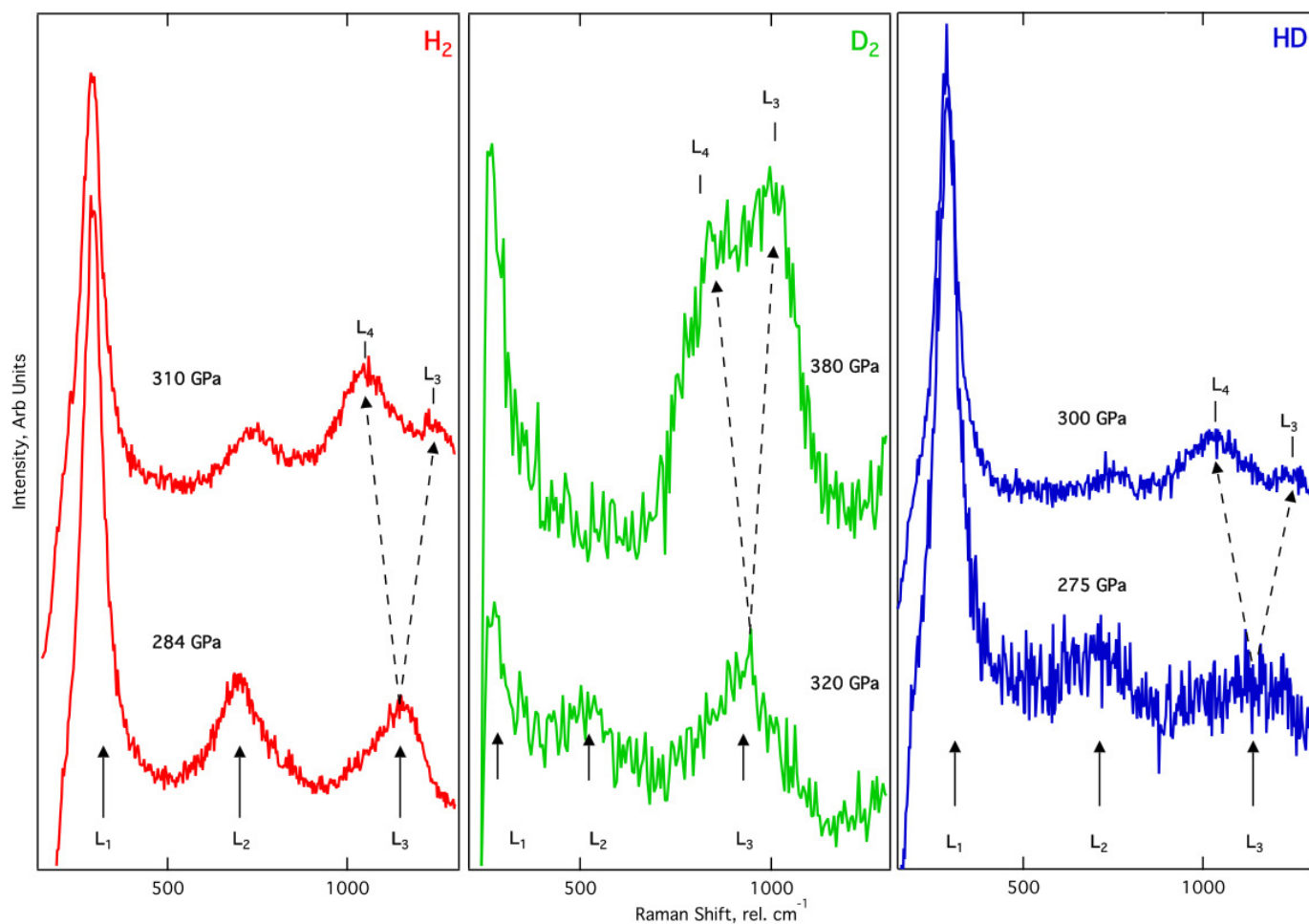
**a**, Vibration-mode (vibron) frequency plotted using the three pressure gauges of the stressed-diamond frequency proposed by Akahama *et al.*: blue squares<sup>27</sup>, green circles<sup>28</sup> and red triangles<sup>11</sup>. **b**, The three pressure

gauges plotted as a function of pressure, coloured as in **a**; the dashed line marks the highest frequency of the stressed diamond recorded on a HD sample, 1,936 rel.  $\text{cm}^{-1}$ .



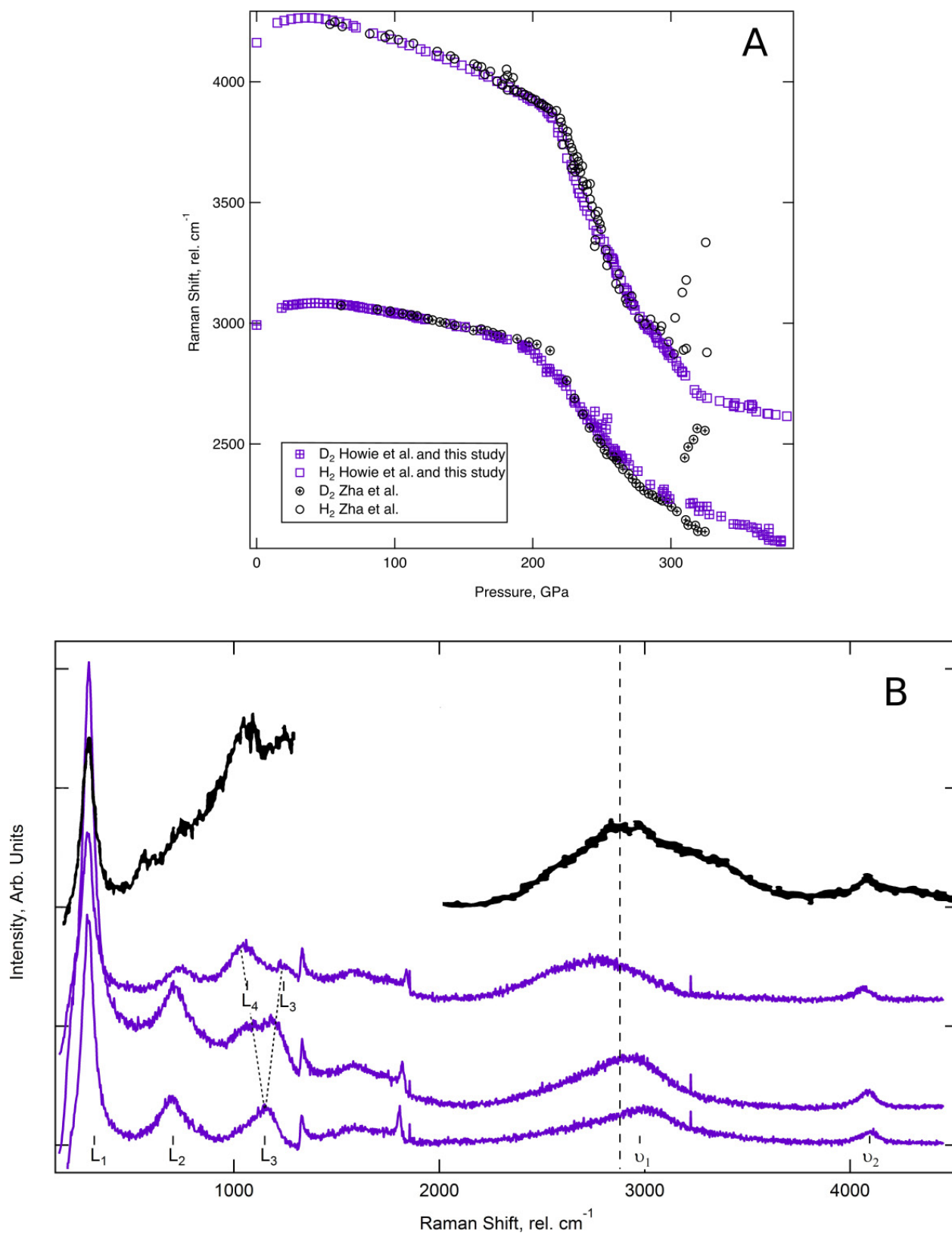
**Extended Data Figure 3 | HD compressed to 218 GPa.** Representative Raman spectra from HD, a mixture of hydrogen (75%) and deuterium (25%), as a function of pressure at 300 K. The spectra show the evolution of the  $\nu_1$  vibrational modes of HD (labelled 'HD- $\nu_1$ ') and  $H_2$  (labelled 'H<sub>2</sub>- $\nu_1$ ') from loading at 0.5–218 GPa, as labelled. Above 47 GPa, there is an observed transfer of integrated intensity from the  $\nu_1$  band of  $H_2$  to the  $\nu_1$  band of HD, with the latter vibrational mode becoming stronger than the former at 150 GPa, and the only resolvable  $\nu_1$  band above 218 GPa. The spectra were collected using a 514-nm excitation wavelength. The spectra from this run above 218 GPa are shown in Fig. 1b.





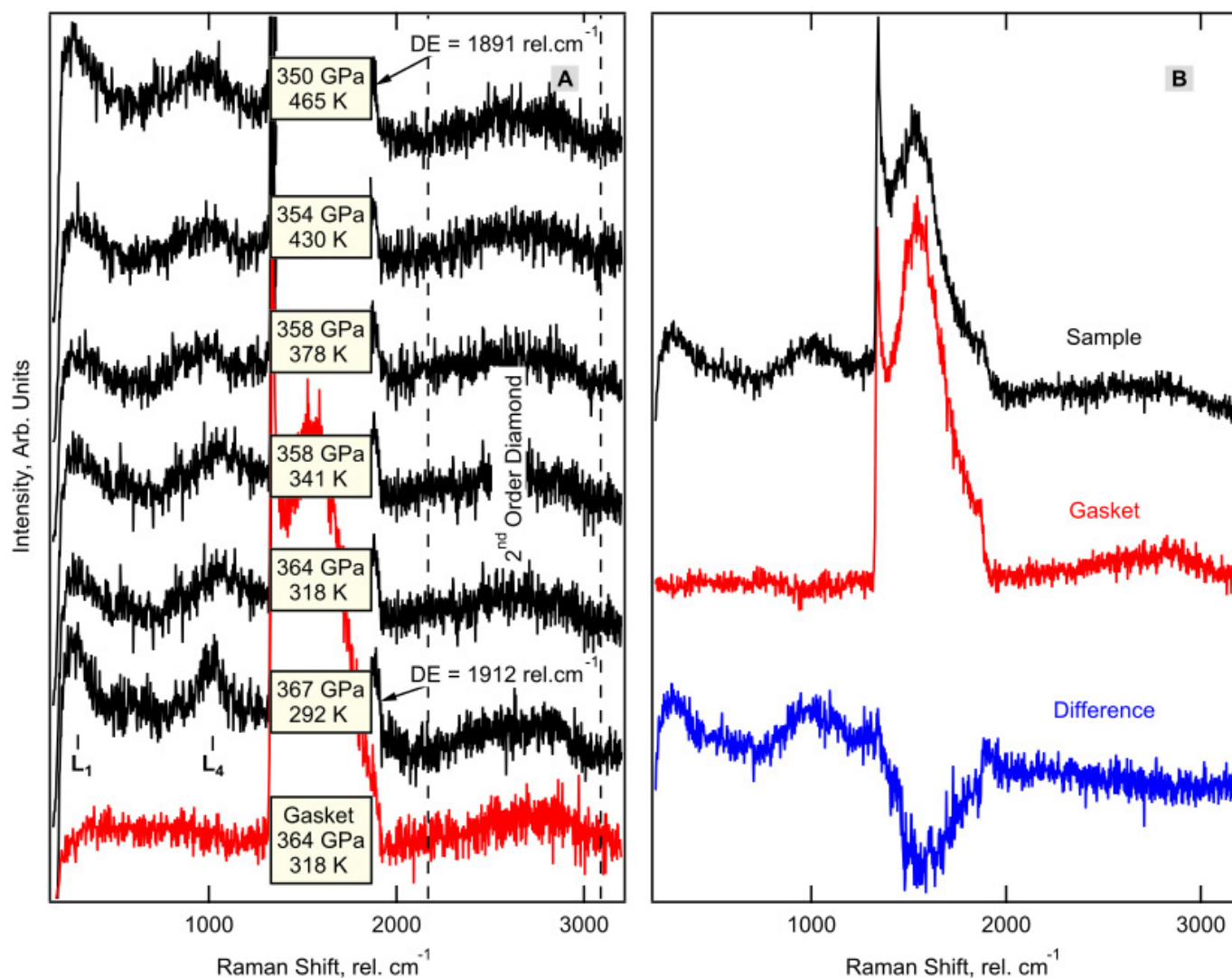
**Extended Data Figure 4 | Low-energy-mode splitting from phase IV to phase IV'.** Representative Raman spectra of the low-frequency excitations of the three isotopes (left, H<sub>2</sub>; centre, D<sub>2</sub>; right, HD) as functions of

pressure during the transition from phase IV to phase IV'. The low-frequency mode *L*<sub>3</sub> splits to produce mode *L*<sub>4</sub>.



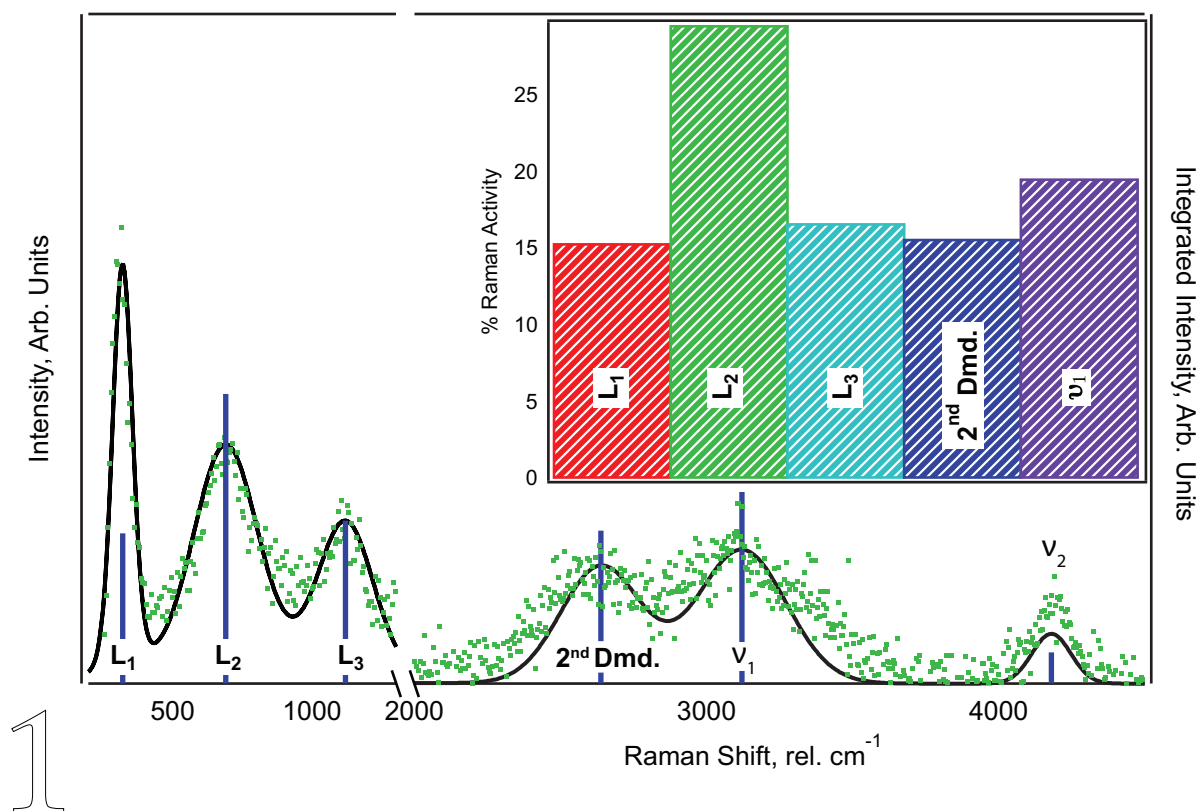
**Extended Data Figure 5 | Comparison with previous data.** **a**, Frequencies of the vibrational modes versus pressure from ref. 21 (black circles) and the current study and our previous study<sup>16</sup> (violet squares). The open symbols represent data for  $\text{H}_2$ ; the symbols enclosing pluses represent

data for  $\text{D}_2$ . **b**, Representative Raman spectra of hydrogen from ref. 21 (black) and ref. 16 (violet). The dashed vertical line indicates the lowest vibrational-mode frequency (and therefore the highest pressure) observed in ref. 21.



**Extended Data Figure 6 | Heating at about 360 GPa.** **a**, Raman spectra for a pure hydrogen sample, taken using a probe laser with a wavelength of 647 nm, as function of temperature at pressures between 367 GPa and 350 GPa (black). The Raman spectrum collected 2  $\mu\text{m}$  away on the rhenium gasket is shown in red. The vertical dashed lines indicate

the frequency space occupied by the second-order diamond band. DE, diamond edge. **b**, Example spectrum of the sample (black) and the gasket (2  $\mu\text{m}$  away, red), collected at 361 GPa, and the difference between them (blue).



**Extended Data Figure 7 | Calculating relative integrated intensities.** Representative Raman spectrum demonstrating how intensities were calculated for a given spectrum. The best fit (black curve) to the experimental data (green points) is shown (measured by the left axis) along with the integrated intensities for each excitation (indicated by

the height of the blue bars, measured by the right axis). The inset shows the raw value of the integrated intensities of each excitation in the main figure as a percentage of the total Raman activity. 2<sup>nd</sup> Dmd., second-order diamond band (unlabelled in the main figure).



# Partially oxidized atomic cobalt layers for carbon dioxide electroreduction to liquid fuel

Shan Gao<sup>1</sup>, Yue Lin<sup>1</sup>, Xingchen Jiao<sup>1</sup>, Yongfu Sun<sup>1,2</sup>, Qiquan Luo<sup>1</sup>, Wenhua Zhang<sup>1</sup>, Dianqi Li<sup>1</sup>, Jinlong Yang<sup>1</sup> & Yi Xie<sup>1,2</sup>

Electroreduction of CO<sub>2</sub> into useful fuels, especially if driven by renewable energy, represents a potentially ‘clean’ strategy for replacing fossil feedstocks and dealing with increasing CO<sub>2</sub> emissions and their adverse effects on climate<sup>1–4</sup>. The critical bottleneck lies in activating CO<sub>2</sub> into the CO<sub>2</sub><sup>•−</sup> radical anion or other intermediates that can be converted further, as the activation usually requires impractically high overpotentials. Recently, electrocatalysts based on oxide-derived metal nanostructures have been shown<sup>5–8</sup> to enable CO<sub>2</sub> reduction at low overpotentials. However, it remains unclear how the electrocatalytic activity of these metals is influenced by their native oxides, mainly because microstructural features such as interfaces and defects<sup>9</sup> influence CO<sub>2</sub> reduction activity yet are difficult to control. To evaluate the role of the two different catalytic sites, here we fabricate two kinds of four-atom-thick layers: pure cobalt metal, and co-existing domains of cobalt metal and cobalt oxide. Cobalt mainly produces formate (HCOO<sup>−</sup>) during CO<sub>2</sub> electroreduction; we find that surface cobalt atoms of the atomically thin layers have higher intrinsic activity and selectivity towards formate production, at lower overpotentials, than do surface cobalt atoms on bulk samples. Partial oxidation of the atomic layers further increases their intrinsic activity, allowing us to realize stable current densities of about 10 milliamperes per square centimetre over 40 hours, with approximately 90 per cent formate selectivity at an overpotential of only 0.24 volts, which outperforms previously reported metal or metal oxide electrodes evaluated under comparable conditions<sup>1,2,6,7,10</sup>. The correct morphology and oxidation state can thus transform a material from one considered nearly non-catalytic for the CO<sub>2</sub> electroreduction reaction into an active catalyst. These findings point to new opportunities for manipulating and improving the CO<sub>2</sub> electroreduction properties of metal systems, especially once the influence of both the atomic-scale structure and the presence of oxide are mechanistically better understood.

To explore the catalytic role of metal sites and metal oxide sites, we first construct a system containing a 4-atom-thick metal layer and then create the corresponding metal oxide on its surface. The atomic thickness of this model system ensures that most of the metal atoms are present as either surface atoms or surface ions<sup>11,12</sup>, so we can explore how the presence of a surface oxide influences the catalytic activity of the corresponding metal. We focus on cobalt (Co) because metals with loosely bonded *d* electrons and the resulting high electrical conductivity are promising for CO<sub>2</sub> reduction<sup>13</sup>, and because both Co and its oxide are widely used catalysts<sup>14–16</sup>. Importantly, the spontaneous oxidation of Co nanostructures in air is relatively slow<sup>17</sup>, and the use of other gases makes it possible to manipulate and control the oxidation process<sup>18</sup>. However, strong in-plane bonds and the lack of an intrinsic driving force for two-dimensional anisotropic growth make the synthesis of Co atomic layers and the controlled conversion of such layers into partially oxidized atomic layers very challenging.

We produced freestanding 4-atom-thick Co sheets with and without surface Co oxide using a ligand-confined growth strategy, in which

the use of dimethylformamide and *n*-butylamine proved crucial for reducing the metal ions and enforcing a sheet-like morphology, respectively (Extended Data Fig. 1). The starting reagent cobalt(III) acetylacetonate, Co(acac)<sub>3</sub>, initially hydrolyses into [Co(H<sub>2</sub>O)<sub>6</sub>]<sup>3+</sup>, on which *n*-butylamine is adsorbed to reduce surface energy and avoid aggregation (Fig. 1a and Extended Data Fig. 2a) until sheet-like products gradually appear during the subsequent condensation process (Extended Data Fig. 2b). Controlled fabrication of either partially oxidized or pure Co atomic layers is achieved by using dimethylformamide to gradually reduce the cobalt ions<sup>19</sup> (Fig. 1a), illustrated by the products obtained at 220 °C after reaction times of either 3 h or 48 h.

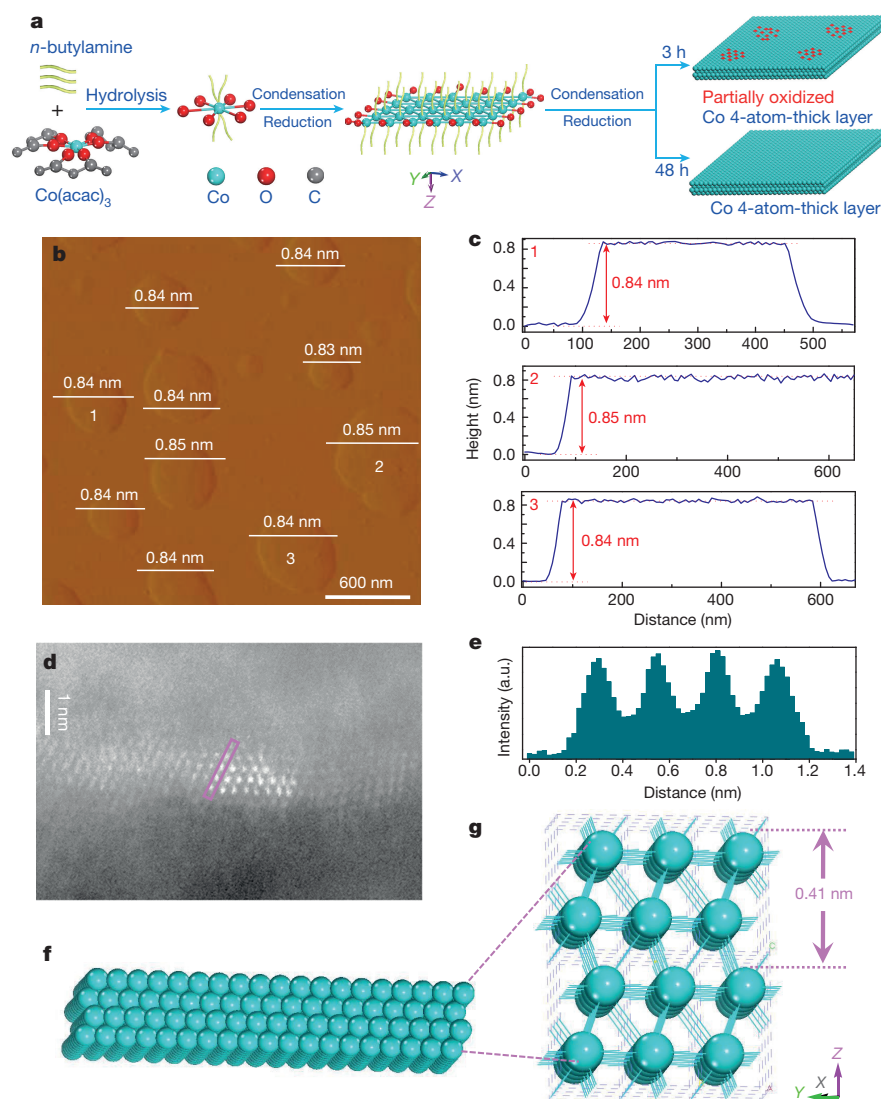
Transmission electron microscope (TEM) images of the product obtained after 3 h reveal a sheet-like morphology, while the powder X-ray diffraction (XRD) pattern can be readily indexed to hexagonal Co (Extended Data Fig. 3a, b). High-resolution TEM images demonstrate that the majority of these two-dimensional sheets correspond to the [001]-oriented hexagonal Co (Fig. 2a, b and Extended Data Fig. 3c), with average sheet thickness of 0.84 nm determined with atomic force microscopy (Fig. 1b, c), close to the 0.82-nm thickness of a 4-atom-thick Co slab along the [001] direction. Lateral high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) imaging confirms the 4-atom layer thickness (Fig. 1d–g).

However, the high-resolution TEM images in Fig. 2a and c also reveal the presence of another distinct structural domain with an interplanar spacing of 0.205 nm and dihedral angle of 90°. This domain corresponds to the (400) plane of cubic Co<sub>3</sub>O<sub>4</sub>, which is embedded in the metallic Co lattice (see schemes in Fig. 2d and e; another high-resolution TEM image from a larger area of an individual sheet showing Co oxide embedded in Co metal is provided in Extended Data Fig. 3c). Elemental mapping (Fig. 2f–h) supports the conclusion that Co metal and Co oxide co-exist in this sample. This is also consistent with the observation of micro-Raman peaks at 482 cm<sup>−1</sup>, 523 cm<sup>−1</sup>, 621 cm<sup>−1</sup> and 694 cm<sup>−1</sup> (Fig. 2i), which correspond to a Co<sub>3</sub>O<sub>4</sub> phase<sup>20</sup> and that disappear upon increasing the reaction time to 48 h. Taken together, these observations demonstrate that the 4-atom-thick Co sheets obtained after 3 h at 220 °C contain some Co oxide, whereas increasing the reaction time to 48 h leads to the formation of pure Co 4-atom-thick layers.

To characterize the performance of the materials as CO<sub>2</sub> reduction electrocatalysts, they were loaded onto a glassy carbon electrode that served as the working electrode; linear sweep voltammetry was carried out using a CO<sub>2</sub>-saturated 0.1 M Na<sub>2</sub>SO<sub>4</sub> solution in a three-electrode set-up. Use of the partially oxidized Co 4-atom-thick layers generates a current density of 10.59 mA cm<sup>−2</sup> at −0.85 V versus a saturated calomel electrode (SCE) (Fig. 3a), roughly 10, 40 and 260 times larger than the current densities obtained with the pure-Co 4-atom-thick layer (Extended Data Fig. 4), the partially oxidized bulk Co and bulk Co (Extended Data Fig. 5), respectively.

Quantification of the solution-phase products by <sup>1</sup>H nuclear magnetic resonance (NMR) shows that of the four samples, the partially

<sup>1</sup>Hefei National Laboratory for Physical Sciences at Microscale, Collaborative Innovation Center of Chemistry for Energy Materials, University of Science and Technology of China, Hefei, Anhui 230026, China. <sup>2</sup>Hefei Science Center of CAS, Hefei, Anhui 230061, China.



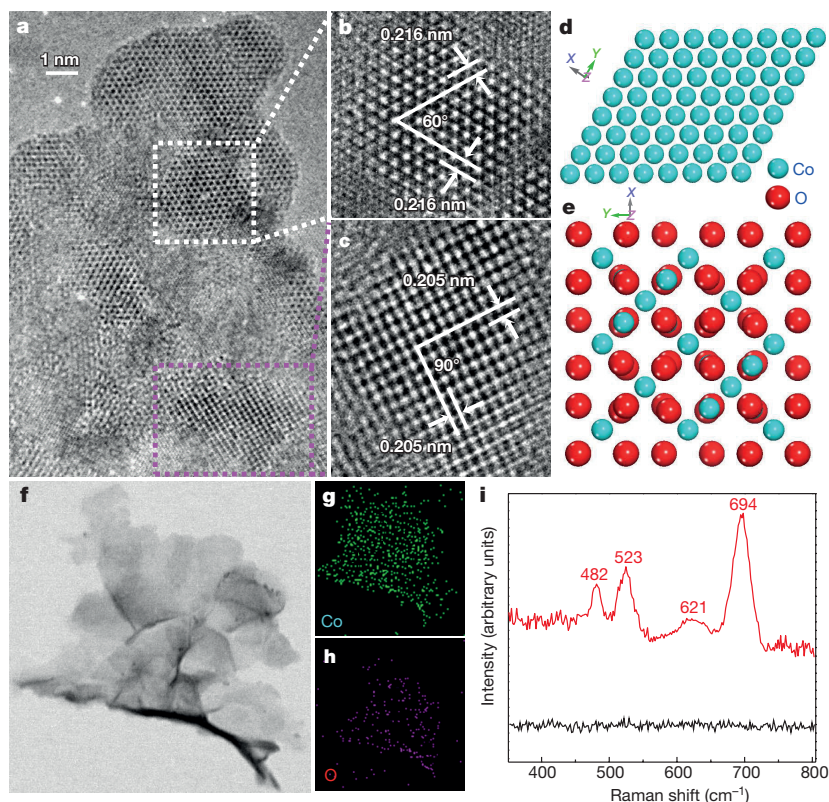
**Figure 1 | Synthetic scheme and characterizations of Co 4-atom-thick layers with and without surface oxide.** **a**, Schematic formation process of the partially oxidized and pure-Co 4-atomic-layer, respectively. **b–g**, Characterizations for the partially oxidized Co 4-atomic-layers: atomic force microscopy image (**b**) and the corresponding height profiles (**c**) (we note that the numbers from 1 to 3 in **c** correspond to the numbers from 1 to 3 in **b**), lateral HAADF-STEM image (**d**) and the corresponding intensity profile along the pink rectangle in **d**, directly showing the 4-atom thickness of the layer (**e**), and the corresponding crystal structures (**f**, **g**). a.u., arbitrary units.

oxidized Co 4-atom-thick layers attain the highest Faradaic efficiency for formate production of 90.1% at  $-0.85$  V versus SCE (Fig. 3b and Extended Data Fig. 6a, b)<sup>21</sup>.  $\text{H}_2$  evolution, quantified by gas chromatography, accounts for the remaining  $\sim 10\%$  of the charges passed. The linear sweep voltammetry curves in  $\text{N}_2$ -saturated  $0.1$  M  $\text{Na}_2\text{SO}_4$  solution indicate that the ultrathin structure and the presence of surface Co oxide also increase the  $\text{H}_2\text{O}$  reduction activity of the catalyst system (Fig. 3a). A  $^{13}\text{CO}_2$  labelling experiment involving the same set-up and 8-h electrolysis yielded a product that generates an obvious  $^{13}\text{C}$  NMR peak at 168.5 parts per million, attributed to  $\text{H}^{13}\text{COO}^-$ , and a  $^1\text{H}$  NMR doublet, corresponding to the proton coupled to the  $^{13}\text{C}$  of  $\text{H}^{13}\text{COO}^-$  (Extended Data Fig. 6c, d)<sup>22</sup>, confirming that formate is indeed derived from  $\text{CO}_2$ .

Taking the reduction potential of  $E = -0.61$  V versus SCE for the  $\text{CO}_2/\text{HCOO}^-$  couple in  $\text{CO}_2$ -saturated  $0.1$  M  $\text{Na}_2\text{SO}_4$  solution<sup>3</sup>, the potential of  $-0.85$  V versus SCE used in our experiments corresponds to an overpotential of  $0.24$  V. To the best of our knowledge, achieving at such a low overpotential a current density as high as  $10.59$  mA  $\text{cm}^{-2}$  and 90.1% formate selectivity has not been possible with any of the previously reported metal or metal oxide electrodes evaluated under comparable conditions<sup>1,2,6,7,10</sup>. Intriguingly, when using the partially oxidized Co 4-atom-thick layers as the working electrode,  $\text{CO}_2$  reduction initiated at  $-0.68$  V versus SCE with a measured Faradaic efficiency for formate formation of 2.3% (Fig. 3a, b). This corresponds to an overpotential of only  $0.07$  V, comparable to that achieved with highly active Pd nanoparticles dispersed on a carbon support<sup>21</sup>.

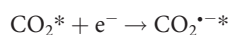
The remarkably enhanced activity of the ultrathin nanostructured catalysts is partly due to their increased electrochemical surface area (ECSA; see Methods for details) that provides a larger number of catalytically active sites. The fivefold ECSA increase from bulk Co to Co 4-atom-thick layers is an important contributor to the 26-fold increase in catalytic activity (Fig. 3a, c). Interestingly, the partially oxidized Co 4-atom-thick layers exhibited nearly the same ECSA as the Co 4-atom-thick layers, yet also a tenfold higher catalytic activity that must therefore be due to the presence of intrinsically more active sites associated with the Co oxide. This conclusion is supported by the relative performance of partially oxidized bulk Co and intact bulk Co (Fig. 3a, c), and by the finding that catalytic current densities gradually increase as the amount of Co oxide in the Co 4-atom-thick layers increases (Extended Data Fig. 7). During 40-h electrocatalysis tests, the partially oxidized Co 4-atom-thick layers show negligible decay in current density while maintaining a formate Faradaic efficiency of approximately 90% (Fig. 3d and Extended Data Fig. 6b). This suggests good stability, which is also confirmed by XRD, Raman and TEM characterization before and after use (Extended Data Fig. 8). Stability testing for Co 4-atom-thick layers and bulk Co also indicates that they do not undergo obvious oxidation or corrosion during long-term electrolysis (Extended Data Figs 9 and 10).

Volumetric  $\text{CO}_2$  adsorption measurement, carried out to explore the reason for the enhanced electrocatalytic activity observed, reveals that the Co 4-atom-thick layer system absorbs more  $\text{CO}_2$  than its bulk counterpart and that partial oxidation increases  $\text{CO}_2$  adsorption capacity

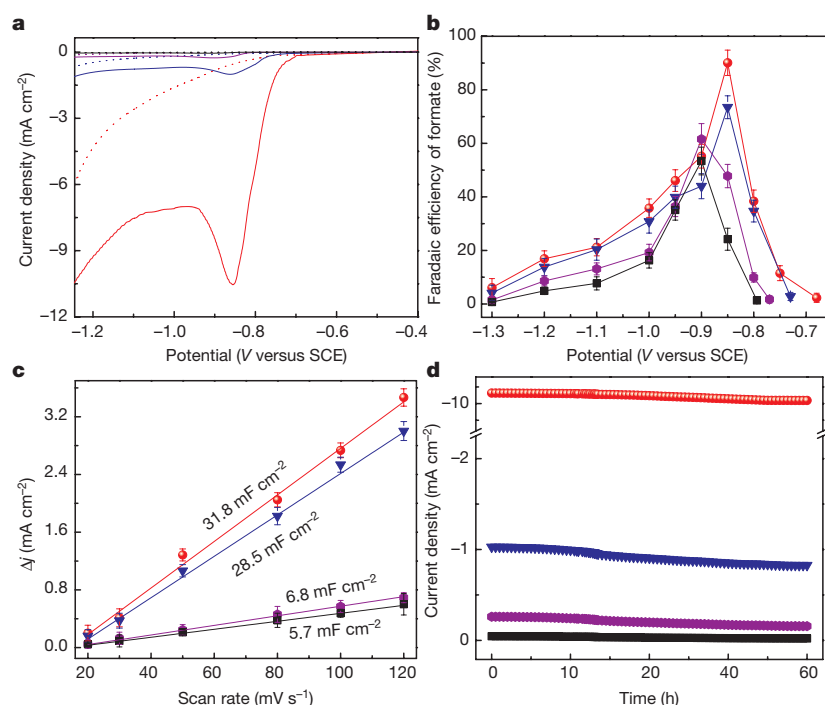


**Figure 2 | Characterizations for the partially oxidized Co 4-atom-thick layers obtained at 220 °C for 3 h.** **a**, High-resolution TEM image. **b**, **c**, Enlarged high-resolution TEM images. **d**, **e**, The related schematic atomic models, clearly showing distinct atomic configuration corresponding to hexagonal Co and cubic  $\text{Co}_3\text{O}_4$ . **f**–**h**, Elemental mapping. **i**, Micro-Raman spectra for the products obtained at 220 °C for 3 h (red line) and 48 h (black line).

further (Fig. 4a). This suggests that the change in oxidation state and increase in surface area synergistically favour  $\text{CO}_2$  adsorption, the prerequisite first event before further reduction reactions can take place. It is commonly accepted that on metal electrodes<sup>3,5–8,23,24</sup>, the adsorbed  $\text{CO}_2$  is initially reduced to the  $\text{CO}_2^{*-}$  intermediate, which could then react further according to:

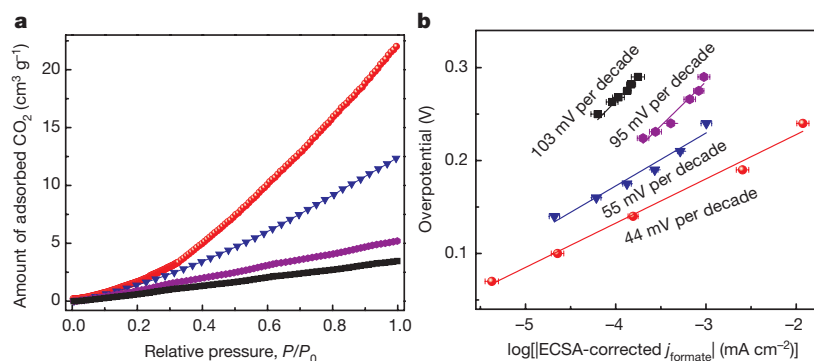


where the asterisk denotes a catalytically active site and  $\text{e}^-$  is an electron. ECSA-corrected Tafel slopes for formate production (see Methods for details) catalysed by the partially oxidized bulk Co and by bulk Co are both close to 118 mV per decade of current (Fig. 4b), indicative of the involvement of a rate-limiting  $1\text{e}^-$  transfer from  $\text{CO}_2$  to  $\text{CO}_2^{*-}$  (refs 5–8).



**Figure 3 | Electroreduction of  $\text{CO}_2$  to formate.** Data are shown for partially oxidized Co 4-atom-thick layers (red), Co 4-atom-thick layers (blue), partially oxidized bulk Co (violet) and bulk Co (black). **a**, Linear sweep voltammetric curves in a  $\text{CO}_2$ -saturated (solid line) and  $\text{N}_2$ -saturated (dashed line) 0.1 M  $\text{Na}_2\text{SO}_4$  aqueous solution. **b**, Faradaic efficiencies of formate at each given potential for 4 h. **c**, Charging current density differences  $\Delta j$  plotted against scan rates. **d**, Chrono-amperometry results at the corresponding potentials (in **b**) with the highest Faradaic efficiencies. The error bars in **b** and **c** represent the standard deviations of five independent measurements of the same sample.





**Figure 4 | Comparison of CO<sub>2</sub> adsorption amount and ECSA-corrected Tafel plots.** Data are shown for the partially oxidized Co 4-atom-thick layers (red), Co 4-atom-thick layers (blue), partially oxidized bulk Co (violet) and bulk Co (black). **a**, CO<sub>2</sub> adsorption isotherms. **b**, ECSA-corrected Tafel plots for formate production.  $j_{\text{formate}}$  is the partial current density of producing formate. The error bars in **b** represent the standard deviations of five independent measurements of the same sample.

In contrast, corresponding Tafel slopes of close to 59 mV per decade obtained with the partially oxidized Co 4-atomic-layers and Co 4-atomic-layers are compatible with a reduction mechanism encompassing a fast pre-equilibrium involving  $1e^-$  transfer to form  $\text{CO}_2^{*-}$  and a subsequent slower chemical reaction as the rate-determining step<sup>5–8</sup>. If this is indeed the case, it appears that Co atoms confined in atomic layers are able to facilitate CO<sub>2</sub> activation by stabilizing the  $\text{CO}_2^{*-}$  intermediate more effectively than can be achieved by their bulk counterpart. We speculate that the further decrease in Tafel slope from 55 mV per decade to 44 mV per decade and the lowering of the onset potential from 0.73 V to 0.68 V upon partial oxidation of the Co 4-atom-thick layers (Figs 3b and 4b) might be due to Co oxide facilitating the rate-determining chemical reaction, probably the  $\text{H}^+$  transfer step (equation (3)). We note, however, that the reaction mechanism remains uncertain and that further efforts are needed to gain in-depth understanding of the individual steps involved.

Our synthetic strategy has allowed us to produce a well controlled model system to explore the influence of both atomic-scale structure and the presence of an oxide on the activity of a metal catalyst. ECSA-corrected Tafel plots and Faradaic efficiencies clearly demonstrate that Co-based catalysts in the form of 4-atom-thick layers exhibit higher intrinsic activity and selectivity for formate production at lower overpotentials than the bulk material, and that partial oxidation improves the intrinsic activity of the system significantly further. Thus the appropriate morphology and oxidation state can transform a material considered nearly non-catalytic for CO<sub>2</sub> reduction into a very active and robust catalyst, calling for a re-thinking of accepted strategies for developing efficient CO<sub>2</sub> electroreduction catalysts.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 March; accepted 10 November 2015.

- Zhang, S., Kang, P. & Meyer, T. J. Nanostructured tin catalysts for selective electrochemical reduction of carbon dioxide to formate. *J. Am. Chem. Soc.* **136**, 1734–1737 (2014).
- Huang, X. F., Cao, T. C., Liu, M. C. & Zhao, G. H. Mg-doped CuFeO<sub>2</sub> photocathodes for photoelectrochemical reduction of carbon dioxide. *J. Phys. Chem. C* **117**, 26432–26440 (2013).
- Rosen, B. A. *et al.* Ionic liquid-mediated selective conversion of CO<sub>2</sub> to CO at low overpotentials. *Science* **334**, 643–644 (2011).
- Zhang, S. *et al.* Polyethylenimine-enhanced electrocatalytic reduction of CO<sub>2</sub> to formate at nitrogen-doped carbon nanomaterials. *J. Am. Chem. Soc.* **136**, 7845–7848 (2014).
- Chen, Y. H., Li, C. W. & Kanan, M. W. Aqueous CO<sub>2</sub> reduction at very low overpotential on oxide-derived Au nanoparticles. *J. Am. Chem. Soc.* **134**, 19969–19972 (2012).
- Li, C. W., Ciston, J. & Kanan, M. W. Electroreduction of carbon monoxide to liquid fuel on oxide-derived nanocrystalline copper. *Nature* **508**, 504–507 (2014).
- Li, C. W. & Kanan, M. W. CO<sub>2</sub> reduction at low overpotential on Cu electrodes resulting from the reduction of thick Cu<sub>2</sub>O films. *J. Am. Chem. Soc.* **134**, 7231–7234 (2012).
- Chen, Y. H. & Kanan, M. W. Tin oxide dependence of the CO<sub>2</sub> reduction efficiency on tin electrodes and enhanced activity for tin/tin oxide thin-film catalysts. *J. Am. Chem. Soc.* **134**, 1986–1989 (2012).

- Sun, Y. F., Gao, S., Lei, F. C., Xiao, C. & Xie, Y. Ultrathin two-dimensional inorganic materials: new opportunities for solid state nanochemistry. *Acc. Chem. Res.* **48**, 3–12 (2015).
- Watkins, J. D. & Bocarsly, A. B. Direct reduction of carbon dioxide to formate in high-gas-capacity ionic liquids at post-transition-metal electrodes. *ChemSusChem* **7**, 284–290 (2014).
- Sun, Y. F. *et al.* Pits confined in ultrathin cerium(IV) oxide for studying catalytic centers in carbon monoxide oxidation. *Nature Commun.* **4**, 2899 (2013).
- Sun, Y. F. *et al.* Fabrication of flexible and freestanding zinc chalcogenide single layers. *Nature Commun.* **3**, 1057 (2012).
- Asadi, M. *et al.* Robust carbon dioxide reduction on molybdenum disulfide edges. *Nature Commun.* **5**, 4470 (2014).
- Xie, X. W., Li, Y., Liu, Z. Q., Haruta, M. & Shen, W. J. Low-temperature oxidation of CO catalysed by Co<sub>3</sub>O<sub>4</sub> nanorods. *Nature* **458**, 746–749 (2009).
- Roy, S. C., Varghese, O. K., Paulose, M. & Grimes, C. A. Toward solar fuels: photocatalytic conversion of carbon dioxide to hydrocarbons. *ACS Nano* **4**, 1259–1278 (2010).
- Sun, Y. F. *et al.* Atomically-thin non-layered cobalt oxide porous sheets for highly efficient oxygen-evolving electrocatalysts. *Chem. Sci.* **5**, 3976–3982 (2014).
- Salavati-Niasari, M., Fereshteh, Z. & Davar, F. Synthesis of cobalt nanoparticles from [bis(2-hydroxyacetophenato)cobalt(II)] by thermal decomposition. *Polyhedron* **28**, 1065–1068 (2009).
- Rojas, T. C. *et al.* Preparation, characterization and thermal evolution of oxygen passivated nanocrystalline cobalt. *J. Mater. Chem.* **9**, 1011–1017 (1999).
- Huang, X. Q. *et al.* Freestanding palladium nanosheets with plasmonic and catalytic properties. *Nature Nanotechnol.* **6**, 28–32 (2011).
- Varghese, B. *et al.* Co<sub>3</sub>O<sub>4</sub> nanostructures with different morphologies and their field-emission properties. *Adv. Funct. Mater.* **17**, 1932–1939 (2007).
- Min, X. Q. & Kanan, M. W. Pd-catalyzed electrohydrogenation of carbon dioxide to formate: high mass activity at low overpotential and identification of the deactivation pathway. *J. Am. Chem. Soc.* **137**, 4701–4708 (2015).
- Tamaki, Y., Morimoto, T., Koike, K. & Ishitani, O. Photocatalytic CO<sub>2</sub> reduction with high turnover frequency and selectivity of formic acid formation using Ru(II) multinuclear complexes. *Proc. Natl Acad. Sci. USA* **109**, 15673–15678 (2012).
- Kumar, B. *et al.* Renewable and metal-free carbon nanofibre catalysts for carbon dioxide reduction. *Nature Commun.* **4**, 2819 (2013).
- Gattrell, M., Gupta, N. & Co, A. A review of the aqueous electrochemical reduction of CO<sub>2</sub> to hydrocarbons at copper. *J. Electroanal. Chem.* **594**, 1–19 (2006).

**Acknowledgements** This work was financially supported by the National Basic Research Program of China (grant number 2015CB932302), the National Nature Science Foundation (grant numbers 21422107, 21331005, 91422303, 21201157 and 11321503), the Program for New Century Excellent Talents in University (grant number NCET-13-0546), the Youth Innovation Promotion Association of CAS (grant number CX2340000100), the Chinese Academy of Science (grant number XDB01020300), the Fundamental Research Funds for the Central Universities (grant number WK2340000063) and the Scientific Research Grant of the Hefei Science Center of CAS (grant numbers 2015HSC-UE006 and 2015HSC-UP015).

**Author Contributions** Y.X., Y.S. and S.G. conceived the idea and co-wrote the paper. Y.S., S.G., Y.L., X.J. and D.L. carried out the sample synthesis, characterization and CO<sub>2</sub> reduction measurement. Y.S., S.G., Q.L., W.Z. and J.Y. discussed the catalytic process. All the authors contributed to the overall scientific interpretation and edited the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.X. (yxie@ustc.edu.cn) and Y.S. (yfsun@ustc.edu.cn).



## METHODS

**Synthesis of partially oxidized Co 4-atom-thick layers.** In a typical procedure, 100 mg  $\text{Co}(\text{acac})_3$  was added into a solution of 20 ml dimethylformamide, 4 ml  $\text{H}_2\text{O}$  and 1 ml *n*-butylamine. After vigorous stirring for 15 min, the mixture was transferred into a 40-ml Teflon-lined autoclave (before this experiment, the autoclave was initially treated with 20 ml *n*-butylamine at 120 °C for 6 h, and then treated with 35 ml  $\text{H}_2\text{O}$  at 120 °C for 12 h), sealed and heated at 220 °C for 3 h. The system was allowed to cool down to room temperature naturally, and the final product was collected by centrifuging the mixture, washed with cyclohexane and absolute ethanol (1:4) many times, and then dried in vacuum for further characterization.

**Synthesis of Co 4-atom-thick layers.** In a typical procedure, 100 mg  $\text{Co}(\text{acac})_3$  was added into a solution of 20 ml dimethylformamide, 4 ml  $\text{H}_2\text{O}$  and 1 ml *n*-butylamine. After vigorous stirring for 15 min, the mixture was transferred into a 40-ml Teflon-lined autoclave (before this experiment, the autoclave was initially treated with 20 ml *n*-butylamine at 120 °C for 6 h, and then treated with 35 ml  $\text{H}_2\text{O}$  at 120 °C for 12 h), sealed and heated at 220 °C for 48 h. The system was allowed to cool down to room temperature naturally, and the final product was collected by centrifuging the mixture, washed with cyclohexane and absolute ethanol (1:4) many times, and then dried in vacuum for further characterization.

**Synthesis of partially oxidized bulk Co.** In a typical procedure, 1.0 g  $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$  was added into a solution of 15 ml ethylene glycol and 15 ml ethylenediamine. After vigorous stirring for 15 min, the mixture was transferred into a 40-ml Teflon-lined autoclave, sealed and heated at 220 °C for 12 h. The system was allowed to cool down to room temperature naturally, and the final product was collected by centrifuging the mixture, washed with cyclohexane and absolute ethanol (1:4) many times, and then dried in vacuum for further characterization.

**Synthesis of bulk Co.** In a typical procedure, 1.0 g  $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$  was added into a solution of 15 ml ethylene glycol and 15 ml ethylenediamine. After vigorous stirring for 15 min, the mixture was transferred into a 40-ml Teflon-lined autoclave, sealed and heated at 220 °C for 24 h. The system was allowed to cool down to room temperature naturally, and the final product was collected by centrifuging the mixture, washed with cyclohexane and absolute ethanol (1:4) many times, and then dried in vacuum for further characterization.

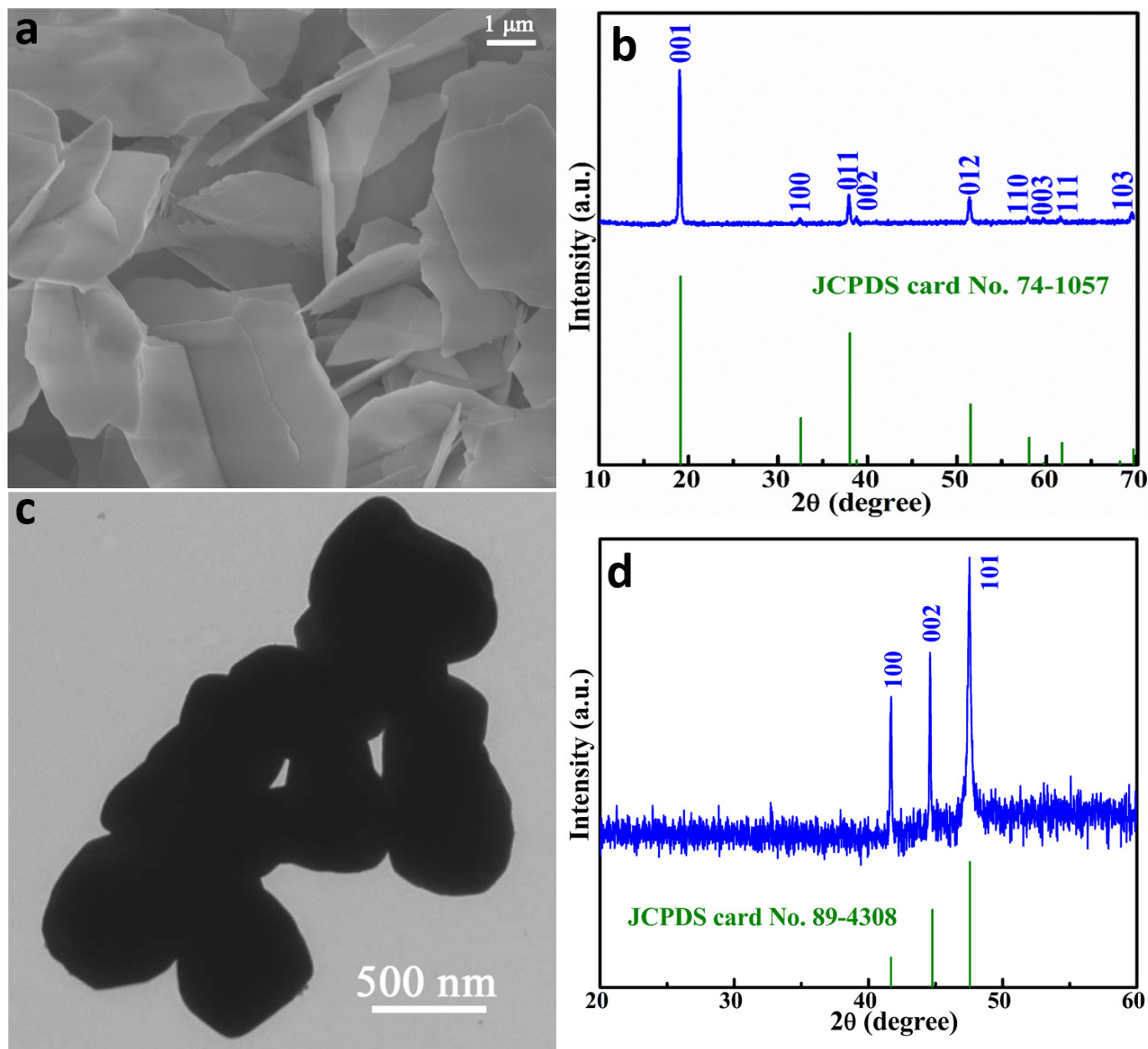
**Characterization.** The field emission scanning electron microscope (SEM) images were performed by using a FEI Sirion-200 SEM. XRD patterns were recorded by using a Philips X'Pert Pro Super diffractometer with  $\text{Cu K}\alpha$  radiation ( $\lambda = 1.54178 \text{ \AA}$ ). Atomic force microscopy in the present work was performed using a Veeco DI Nano-scope MultiMode V system. The TEM was carried out on a JEM-2100F field emission electron microscope at an acceleration voltage of 200 kV. The high-resolution TEM, HAADF-STEM and the corresponding energy dispersive spectroscopy mapping analyses were performed on a JEOL JEM-ARM200F TEM/STEM with a spherical aberration corrector. Raman spectra were detected on a Renishaw RM3000 Micro-Raman system.

**Electrochemical measurements.** Electrochemical measurements were carried out in a three-electrode system at an electrochemical station (CHI760E). Typically,

a 10-mg sample and 40  $\mu\text{l}$  of Nafion solution (5 wt%) were dispersed in 1 ml of water-ethanol solution with a volume ratio of 3:1 by sonicating for 1 h to form a homogeneous ink. Then, 30  $\mu\text{l}$  of the dispersion was loaded onto a glassy carbon electrode with diameter 12 mm. For  $\text{CO}_2$  reduction experiments, linear sweep voltammetry with a scan rate of  $20 \text{ mV s}^{-1}$  was conducted in  $\text{CO}_2$ -saturated 0.1 M  $\text{Na}_2\text{SO}_4$  solution (60 ml,  $\text{pH} \approx 6$ ) (the  $\text{Na}_2\text{SO}_4$  electrolyte was purged with  $\text{CO}_2$  for 30 min before the measurement). For comparison, linear sweep voltammetry with a scan rate of  $20 \text{ mV s}^{-1}$  was also conducted in  $\text{N}_2$ -saturated 0.1 M  $\text{Na}_2\text{SO}_4$  solution. The glassy carbon electrode served as the working electrode. The counter and the reference electrodes were the platinum gauze and the SCE reference electrode, respectively. The liquid products were quantified by NMR (Bruker AVANCE AV III 400) spectroscopy, in which 0.5 ml electrolyte was mixed with 0.1 ml  $\text{D}_2\text{O}$  (deuterated water) and 0.05  $\mu\text{l}$  dimethyl sulfoxide (DMSO, Sigma, 99.99%) was added as an internal standard. The one-dimensional  $^1\text{H}$  spectrum was measured with water suppression using a pre-saturation method. The evolved gas products were detected using an Agilent Technologies 7890B gas chromatograph.

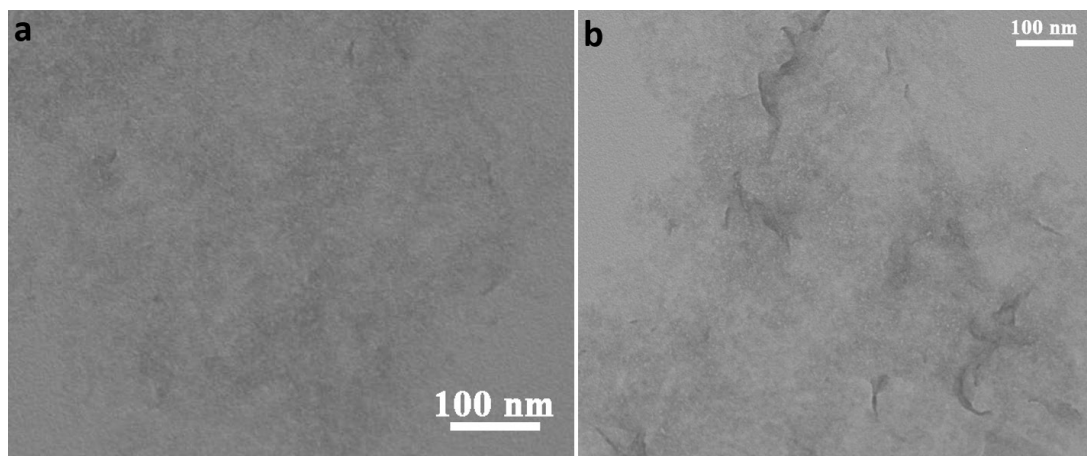
$\text{ECSA} = R_f S$ , in which  $S$  stands for the real surface area of the smooth metal electrode, which was generally equal to the geometric area of glassy carbon electrode (in this work,  $S = 1.13 \text{ cm}^2$ ). The roughness factor  $R_f$  was estimated from the ratio of double-layer capacitance  $C_{\text{dl}}$  for the working electrode and the corresponding smooth metal electrode (assuming that the average double-layer capacitance of a smooth metal surface is  $20 \mu\text{F cm}^{-2}$ )<sup>25</sup>, that is,  $R_f = C_{\text{dl}}/20 \mu\text{F cm}^{-2}$ . The  $C_{\text{dl}}$  was determined by measuring the capacitive current associated with double-layer charging from the scan-rate dependence of cyclic voltammetric stripping. For this, the potential window of cyclic voltammetric stripping was  $-0.3 \text{ V}$  to  $-0.2 \text{ V}$  versus SCE (0.1 M  $\text{Na}_2\text{SO}_4$  solution). The scan rates were  $20 \text{ mV s}^{-1}$ ,  $30 \text{ mV s}^{-1}$ ,  $50 \text{ mV s}^{-1}$ ,  $80 \text{ mV s}^{-1}$ ,  $100 \text{ mV s}^{-1}$  and  $120 \text{ mV s}^{-1}$ . The  $C_{\text{dl}}$  was estimated by plotting the  $\Delta j = (j_a - j_c)$  at  $-0.25 \text{ V}$  (where  $j_c$  and  $j_a$  are the cathodic and anodic current densities, respectively) versus SCE against the scan rate, in which the slope was twice that of  $C_{\text{dl}}$ . ECSA-corrected Tafel slopes for formate production (that is,  $j_{\text{total}} \times \eta_{\text{formate}}/\text{ECSA}$ ) were calculated from the corresponding ECSA-corrected current densities for formate according to the linear sweep voltammetry curves and the formate Faradaic efficiency ( $\eta_{\text{formate}}$ ). The Faradaic efficiency of formate was calculated from the total amount of charge  $Q$  (in units of coulombs) passed through the sample and the total amount of formate produced  $n_{\text{formate}}$  (in moles).  $Q = I \times t$ , where  $I$  (in amperes) is the reduction current at a specific applied potential and  $t$  is the time (in seconds) for the constant reduction current. The total amount of formate produced was measured using NMR (Bruker AVANCE AV III 400) spectroscopy. Assuming that two electrons are needed to produce one formate molecule, the Faradaic efficiency can be calculated as follows: Faradaic efficiency  $= 2F \times n_{\text{formate}}/Q = 2F \times n_{\text{formate}}/(I \times t)$ , where  $F$  is the Faraday constant.

25. Popczyk, M., Serek, A & Budniok, A. Production and properties of composite layers based on an Ni-P amorphous matrix. *Nanotechnology* **14**, 341–346 (2003).

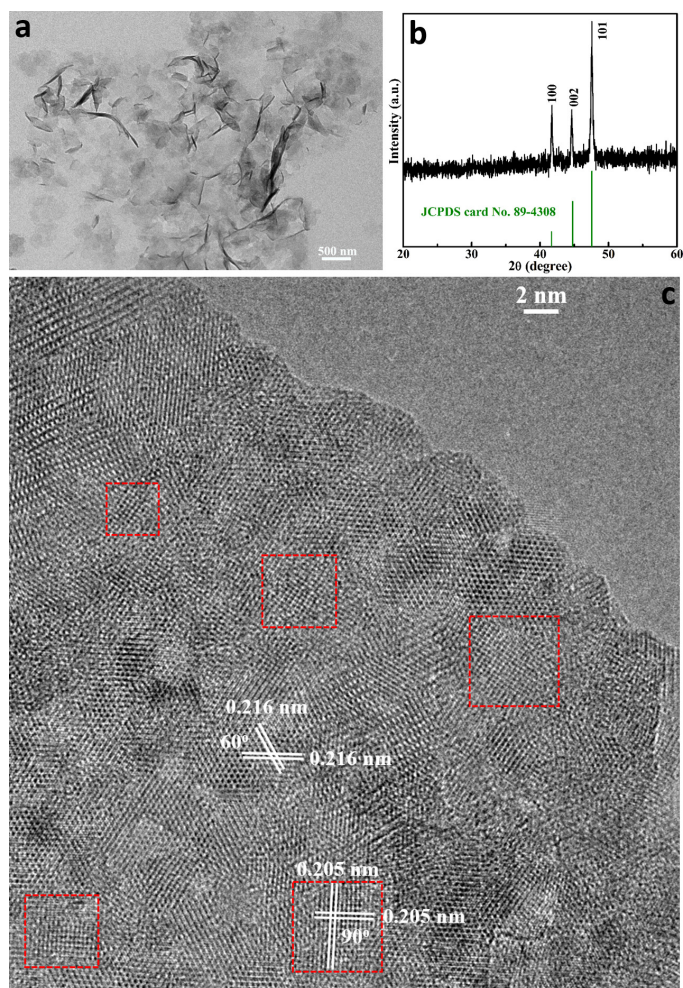


**Extended Data Figure 1 | Characterizations for the comparable products.** **a, b**, SEM image (**a**) and XRD pattern (**b**) for Co(OH)<sub>2</sub> sheets. **c, d**, TEM image (**c**) and XRD pattern (**d**) for large and irregular Co particles. In the case where only *n*-butylamine was present, the reaction produced two-dimensional Co(OH)<sub>2</sub> sheets (**a, b**), whereas the reaction

yielded large and irregular Co particles when only dimethylformamide was used (**c, d**). These results indicated that *n*-butylamine favoured the formation of a sheet-like morphology, while dimethylformamide was beneficial in reducing the cobalt ions with high oxidation states. JCDPS, the Joint Committee on Powder Diffraction Standards.

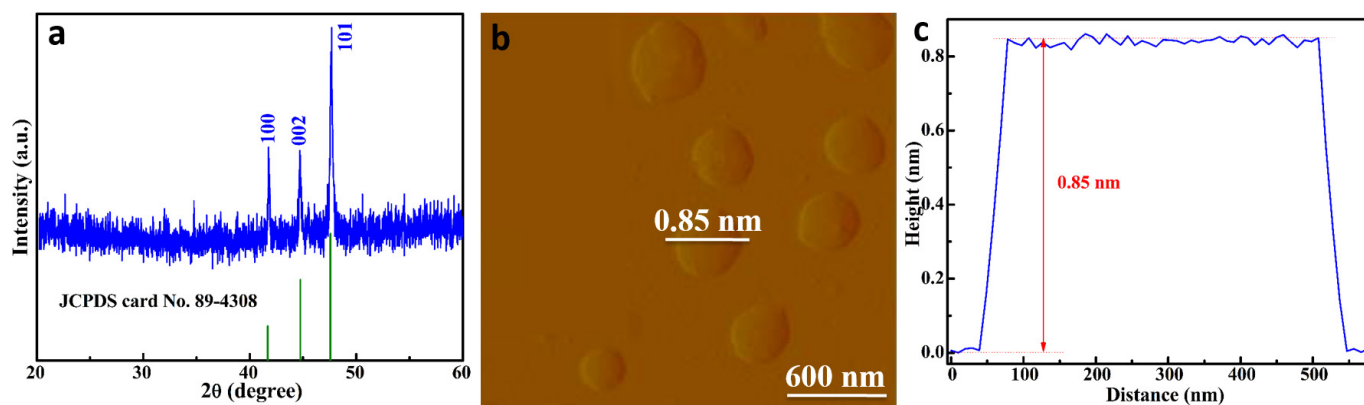


**Extended Data Figure 2 | Characterizations for the intermediate products.** TEM images for the obtained products at 220 °C for 0.5 h (a) and 2 h (b).

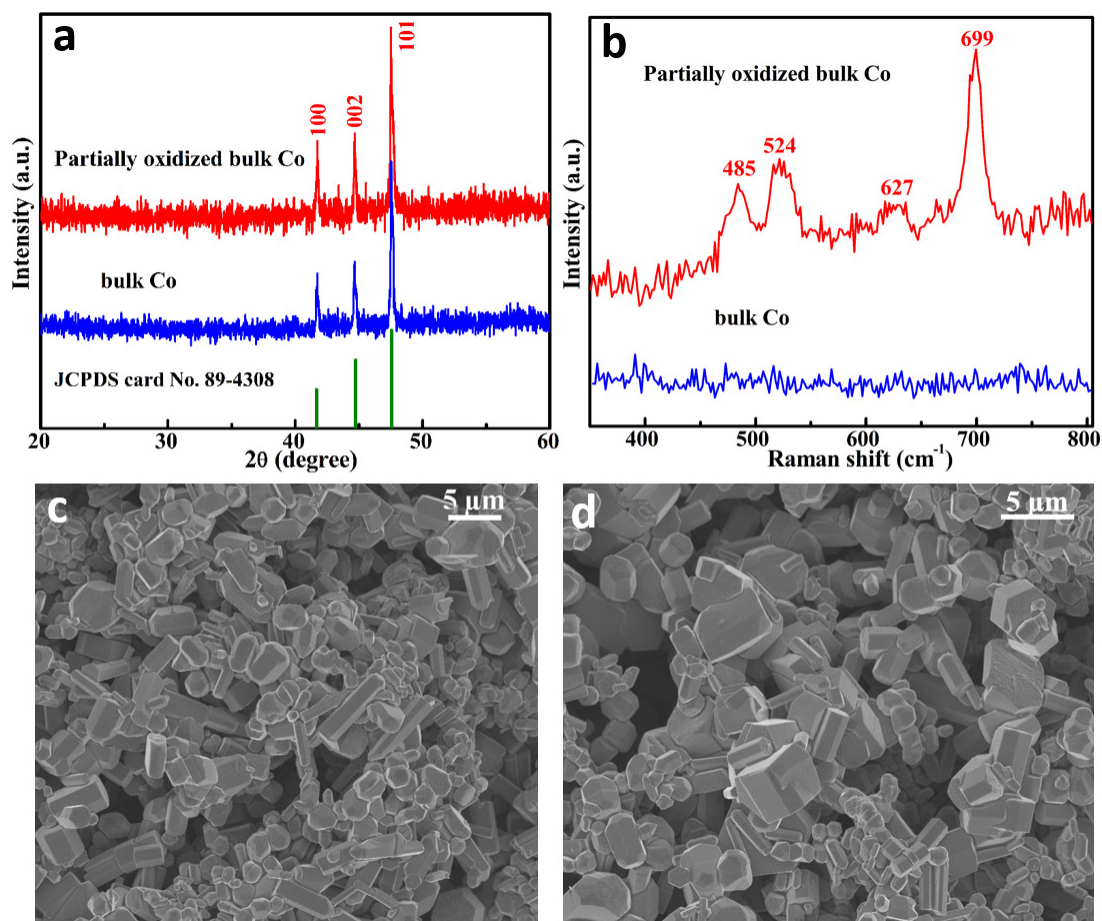


**Extended Data Figure 3 | Supplementary characterizations for the partially oxidized Co 4-atom-thick layers.** **a**, TEM image. **b**, XRD pattern. **c**, High-resolution TEM image, in which the majority of these two-dimensional sheets corresponds to the [001]-oriented hexagonal Co, while the other structural domains denoted by red squares correspond to the cubic  $\text{Co}_3\text{O}_4$ .

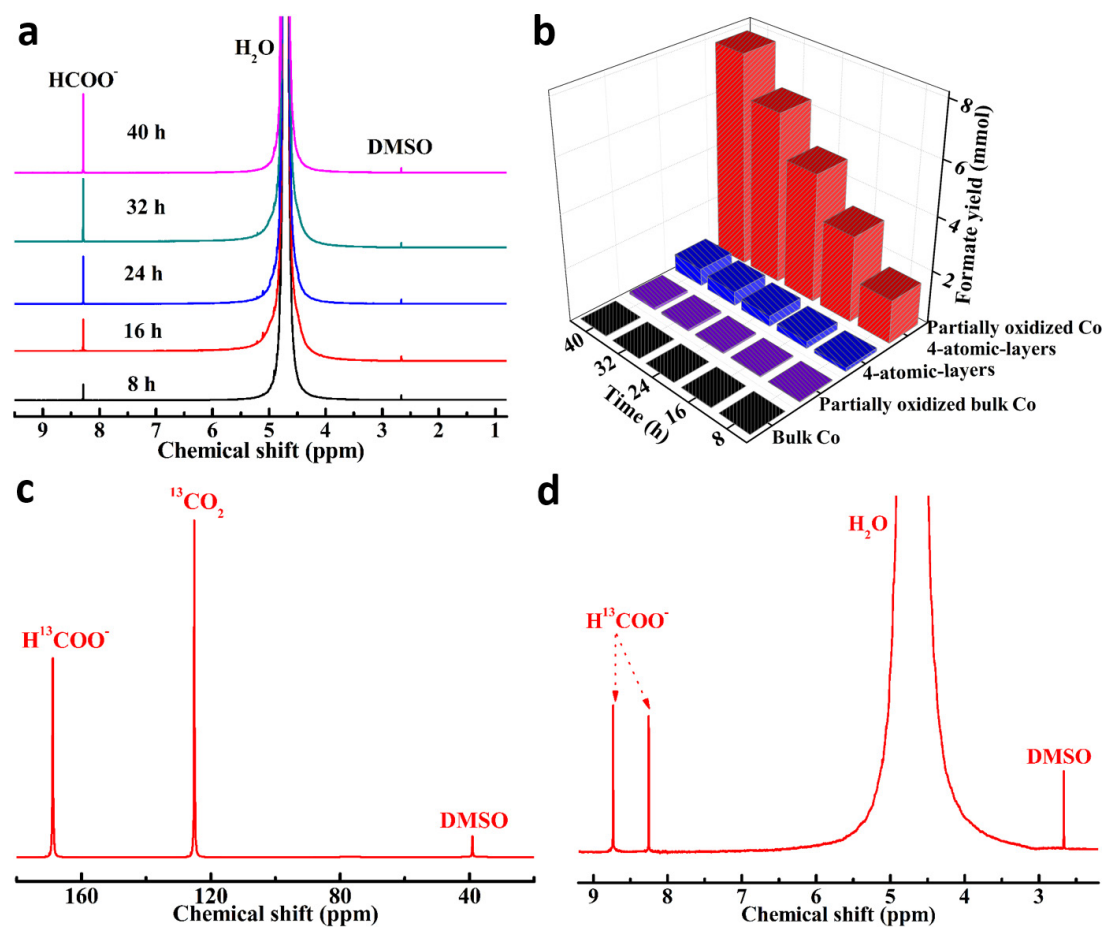




Extended Data Figure 4 | Characterizations for the Co 4-atom-thick layers. **a**, XRD pattern. **b**, Atomic force microscope image. **c**, The corresponding height profile. Data are shown for the products obtained at 220 °C for 48 h.



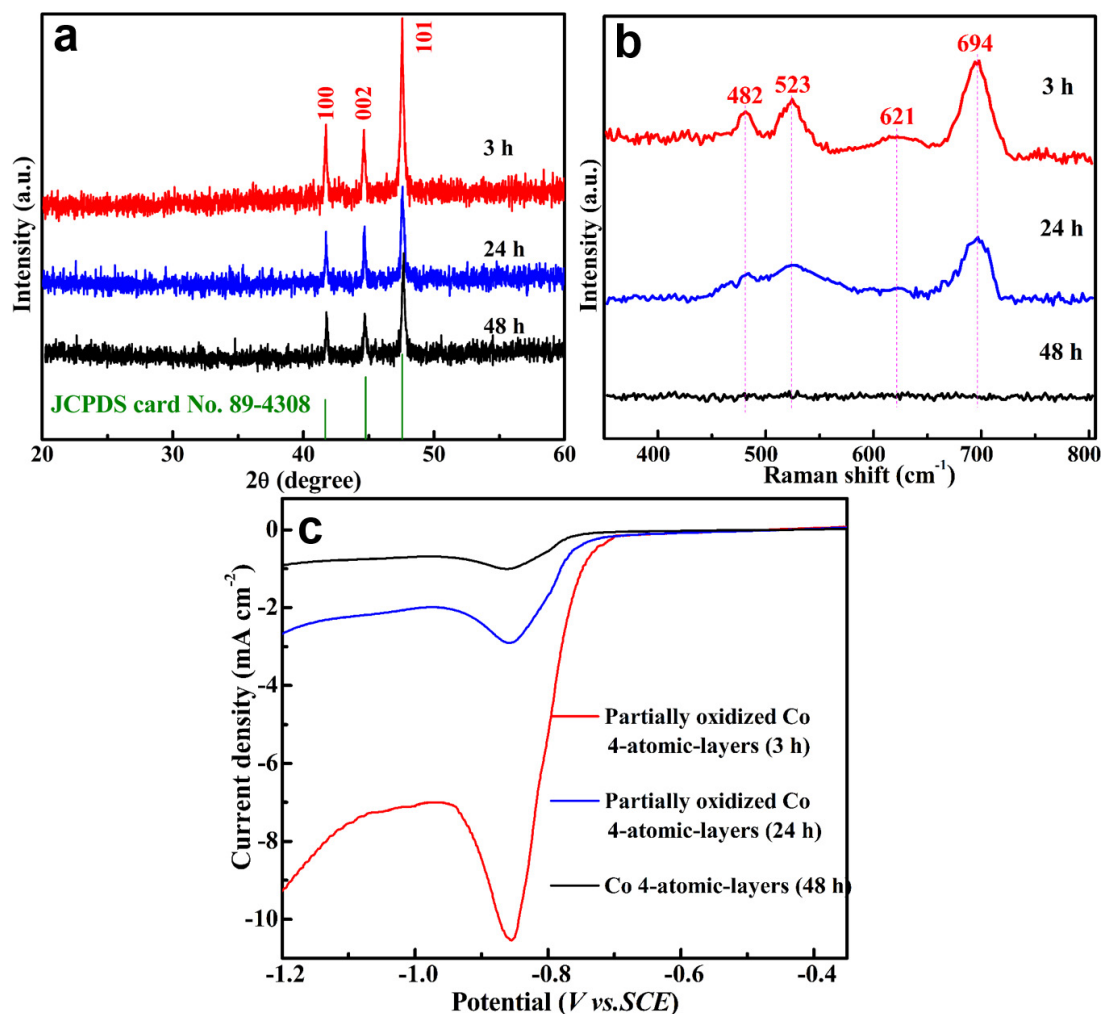
**Extended Data Figure 5 | Characterizations for partially oxidized bulk Co and bulk Co particles. a, XRD patterns. b, Micro-Raman spectra. c, SEM image for partially oxidized bulk Co particles. d, SEM image for bulk Co particles.**



#### Extended Data Figure 6 | NMR spectra and formate yield.

**a**, Representative NMR spectra of the electrolyte after  $\text{CO}_2$  reduction electrolysis at  $-0.85$  V versus SCE for the partially oxidized Co 4-atom-thick layers. DMSO is used as an internal standard for quantification of  $\text{HCOO}^-$ . **b**, Formate yield at the corresponding potentials with the highest Faradaic efficiencies for the partially oxidized Co 4-atom-thick layers,

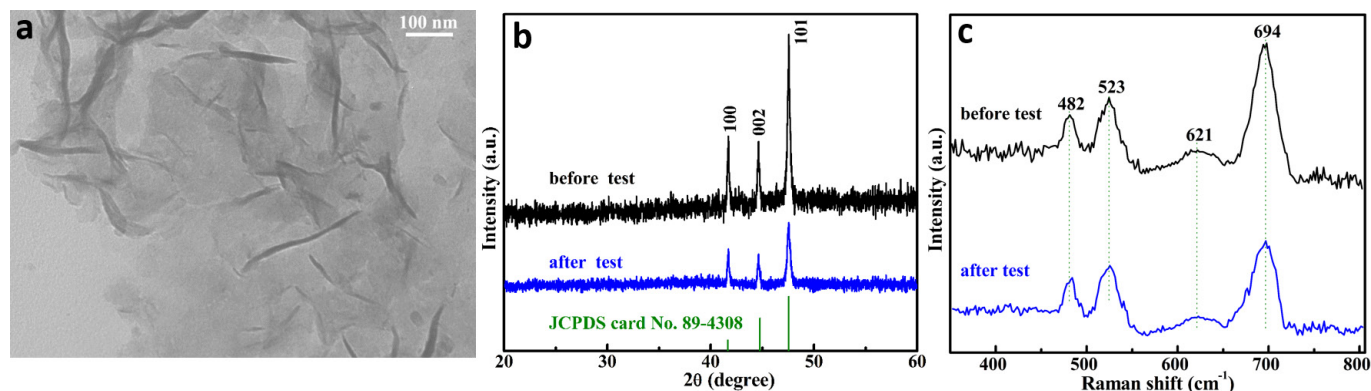
pure Co 4-atom-thick layers, partially oxidized bulk Co and bulk Co. Independently prepared electrodes evaluated under identical conditions in **b** exhibited a variability of  $<10\%$  for the formate yield. **c**, **d**,  $^{13}\text{C}$  spectra (**c**) and  $^1\text{H}$ -NMR spectra (**d**) of the electrolyte after 8 h  $^{13}\text{CO}_2$  reduction electrolysis at  $-0.85$  V versus SCE for the partially oxidized Co 4-atom-thick layers.



**Extended Data Figure 7 | Comparison for Co 4-atom-thick layers in the absence or presence of cobalt oxide with different concentrations.** **a**, XRD patterns. **b**, Raman spectra. **c**, Linear sweep voltammetric curves. Data are for a CO<sub>2</sub>-saturated 0.1 M Na<sub>2</sub>SO<sub>4</sub> aqueous solution for the partially oxidized Co 4-atom-thick layers obtained at 220 °C for 3 h, partially oxidized Co 4-atom-thick layers obtained at 220 °C for 24 h

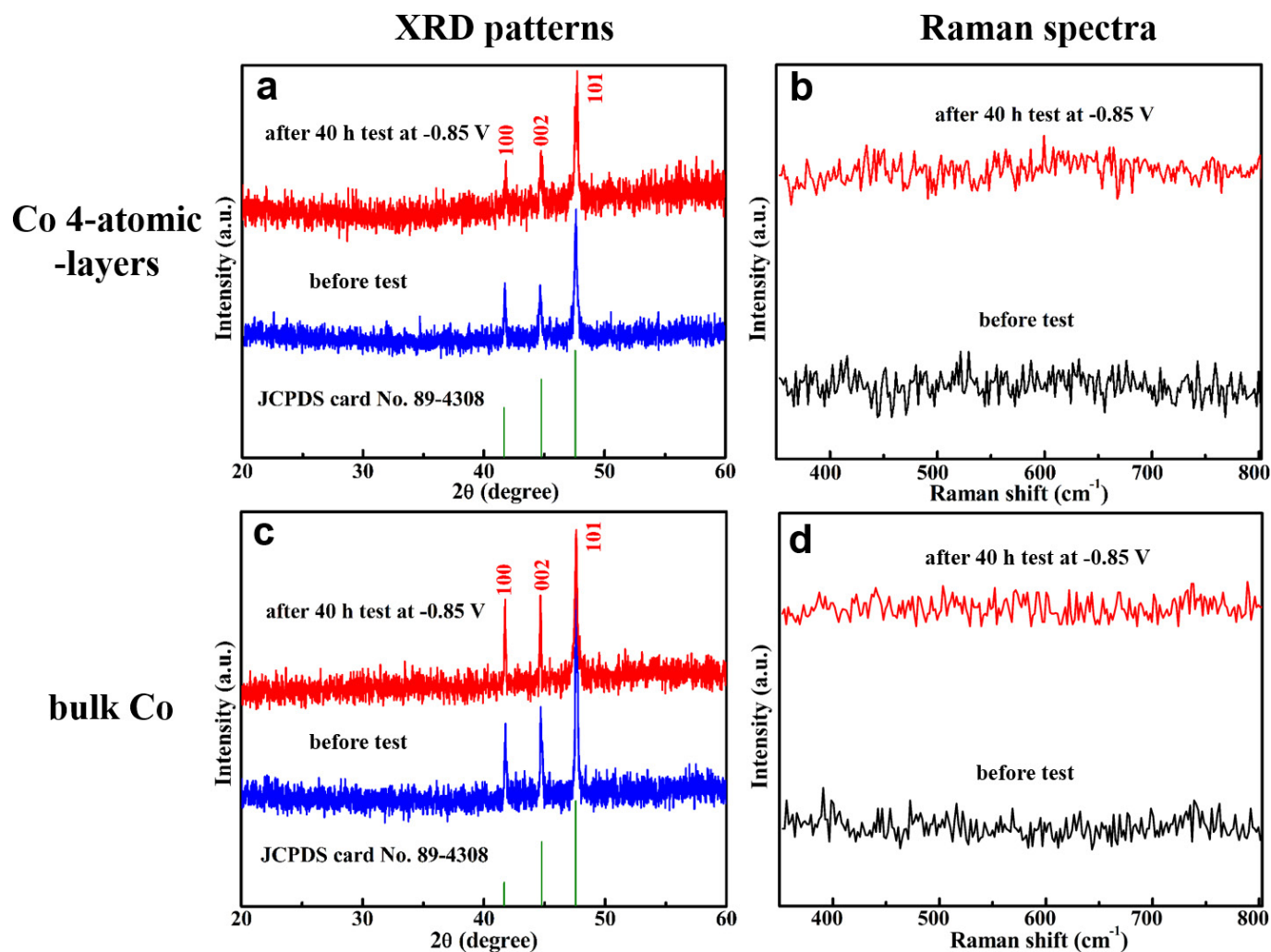
(the synthesis process is the same as that for fabricating the partially oxidized Co 4-atom-thick layers obtained at 220 °C for 3 h except that the synthetic time is increased from 3 h to 24 h; note that the increased reaction time results in the decreased amount of cobalt oxide in the Co 4-atom-thick layers), and pure Co 4-atom-thick layers obtained at 220 °C for 48 h.





**Extended Data Figure 8 | Characterizations for the partially oxidized Co 4-atom-thick layers after the 40-h test.** **a**, TEM image for the partially oxidized Co 4-atom-thick layers after the 40-h  $\text{CO}_2$  reduction test. **b**, **c**, XRD patterns (**b**) and Raman spectra (**c**) for the partially oxidized Co 4-atom-thick layers before and after the 40-h  $\text{CO}_2$  reduction test. The samples for the above characterizations were collected as follows:

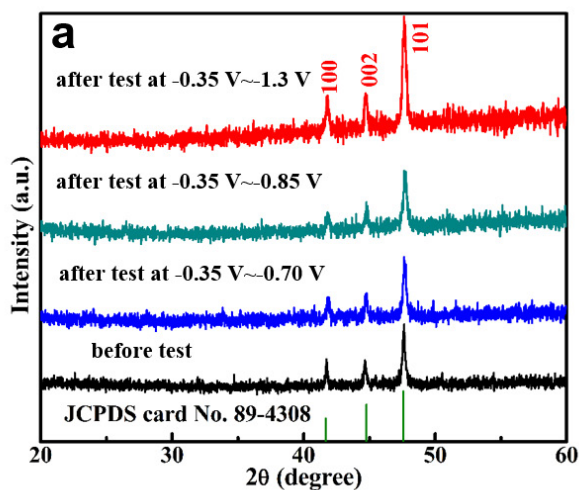
the working electrodes after 40 h of electrolysis were sonicated in ethanol for about 3 min and then the samples were collected by centrifuging the mixture, washed with cyclohexane and absolute ethanol (1:4) many times, and then dried in vacuum. The above process was performed on approximately 50 similar working electrodes and all the samples collected were used to conduct the above characterizations.



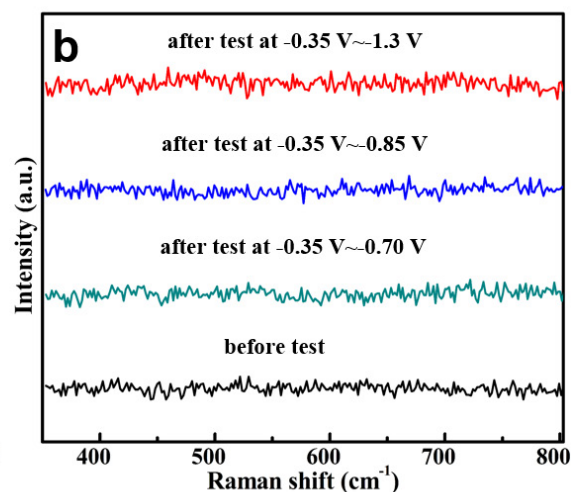
Extended Data Figure 9 | XRD patterns and Raman spectra before and after 40-h electrolysis at  $-0.85$  V versus SCE for Co 4-atom-thick layers and bulk Co. **a, b**, XRD patterns (**a**) and Raman spectra (**b**) for Co 4-atom-thick layers. **c, d**, XRD patterns (**c**) and Raman spectra (**d**) for bulk Co.

Co 4-atomic  
-layers

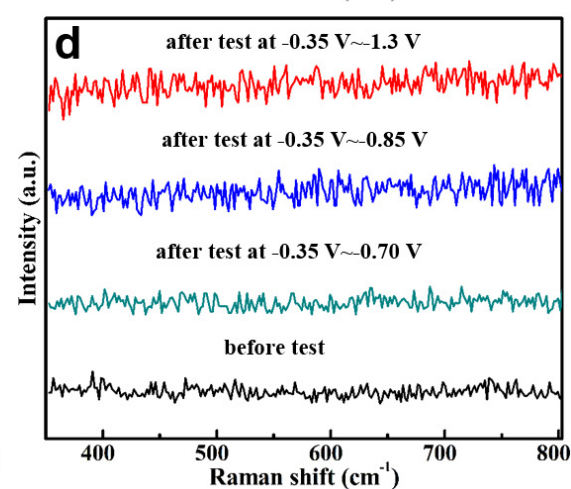
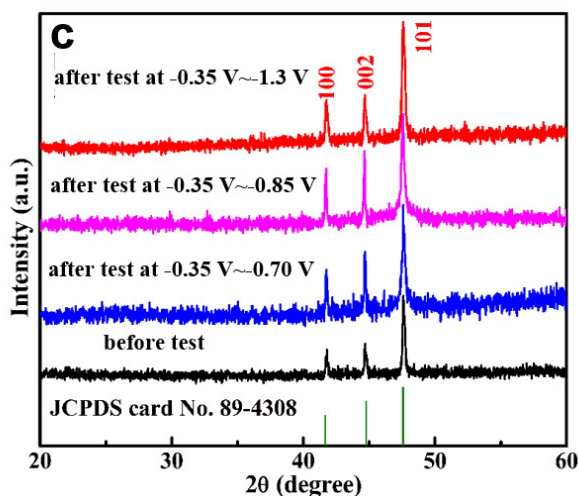
## XRD patterns



## Raman spectra



## bulk Co



Extended Data Figure 10 | XRD patterns and Raman spectra before and after repeating linear sweep voltammetry measurement scanning from  $-0.35$  V versus SCE to different potentials (versus SCE) about 300 times. a, b, XRD patterns (a) and Raman spectra (b) for Co 4-atom-thick layers. c, d, XRD patterns (c) and Raman spectra (d) for bulk Co.

# Four-electron deoxygenative reductive coupling of carbon monoxide at a single metal site

Joshua A. Buss<sup>1</sup> & Theodor Agapie<sup>1</sup>

Carbon dioxide is the ultimate source of the fossil fuels that are both central to modern life and problematic: their use increases atmospheric levels of greenhouse gases, and their availability is geopolitically constrained<sup>1</sup>. Using carbon dioxide as a feedstock to produce synthetic fuels might, in principle, alleviate these concerns. Although many homogeneous and heterogeneous catalysts convert carbon dioxide to carbon monoxide<sup>2</sup>, further deoxygenative coupling of carbon monoxide to generate useful multicarbon products is challenging<sup>3</sup>. Molybdenum and vanadium nitrogenases are capable of converting carbon monoxide into hydrocarbons under mild conditions, using discrete electron and proton sources<sup>4</sup>. Electrocatalytic reduction of carbon monoxide on copper catalysts<sup>5</sup> also uses a combination of electrons and protons, while the industrial Fischer–Tropsch process uses dihydrogen as a combined source of electrons and electrophiles for carbon monoxide coupling at high temperatures and pressures<sup>6</sup>. However, these enzymatic and heterogeneous systems are difficult to probe mechanistically. Molecular catalysts have been studied extensively<sup>6–23</sup> to investigate the elementary steps by which carbon monoxide is deoxygenated and coupled, but a single metal site that can efficiently induce the required scission of carbon–oxygen bonds and generate carbon–carbon bonds has not yet been documented. Here we describe a molybdenum compound, supported by a terphenyl–diphosphine ligand, that activates and cleaves the strong carbon–oxygen bond of carbon monoxide, enacts carbon–carbon coupling, and spontaneously dissociates the resulting fragment. This complex four-electron transformation is enabled by the terphenyl–diphosphine ligand<sup>24,25</sup>, which acts as an electron reservoir and exhibits the coordinative flexibility needed to stabilize the different intermediates involved in the overall reaction sequence. We anticipate that these design elements might help in the development of efficient catalysts for converting carbon monoxide to chemical fuels, and should prove useful in the broader context of performing complex multi-electron transformations at a single metal site.

When using molecular systems to study the conversion of carbon monoxide (CO) into multicarbon products, a common strategy invokes the stepwise addition of hydrides and electrophiles to metal–carbonyl compounds<sup>6</sup>. For example, sequential Lewis-acid-assisted metal–formyl generation, activation and migratory insertion has been demonstrated<sup>6</sup>, realizing a complex reductive coupling sequence in a single step. Under very reducing conditions, highly oxophilic metal centres have been shown to reduce, couple and deoxygenate CO, processes driven by the strength of the metal–oxygen interaction<sup>7–11</sup>. A similar approach involves the insertion of CO into early transition metal hydride bonds; however, the stability imparted by the oxophilic metal centres complicates product release<sup>12–16</sup>. The addition of silyl electrophiles to anionic dicarbonyl complexes of group 5 metals has been shown to generate  $\eta^2$ -bis(siloxyacetylene) compounds<sup>17</sup>, constructing an acetylene diolate linkage similar to that seen for the reduction of CO by alkali metals<sup>18</sup>. Examples of CO coupling are also known for the later transition metals; an anionic iron–dicarbonyl and

a rhodium–porphyrin dimer demonstrate formal reduction but not deoxygenation of CO moieties<sup>19,20</sup>.

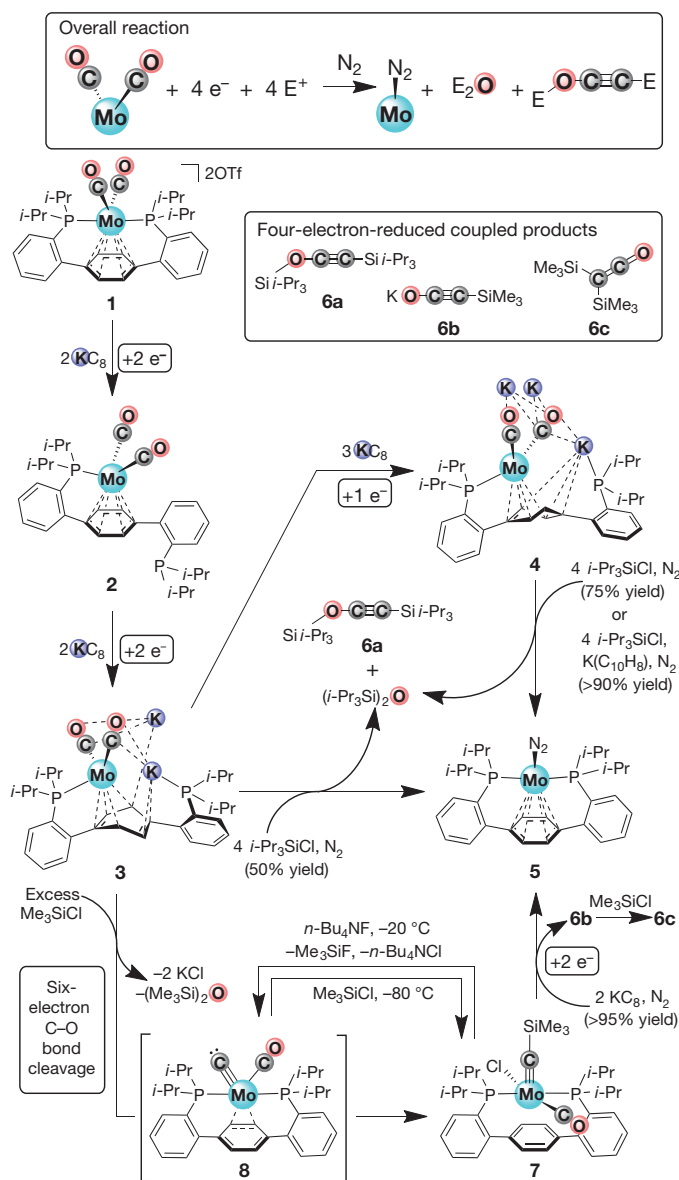
Complete bond-breaking transformations of diatomic small-molecule substrates at a single metal centre present a considerable synthetic challenge, as four or more electrons are often required (for example, N<sub>2</sub> cleavage requires six electrons; O<sub>2</sub>, four electrons; CO, six electrons). Taking this into consideration, the nature of the envisioned intermediates in a deoxygenative CO coupling scheme varies markedly. Before the cleavage of C–O bonds can occur, reduced, electron-rich metal centres that are capable of CO activation are necessary. The scission of CO at a single transition-metal site in the presence of strong reductants has been reported, and leads to the generation of an anionic molybdenum carbide<sup>21</sup>. Following cleavage, the products and potential coupling precursors are more-oxidized complexes bearing metal–ligand multiple bonds. The formation of C–C bonds between CO ligands and substituted alkylidyne moieties is known, but these carbyne ligands do not originate from CO in a direct fashion (refs 22, 23).

Here, to facilitate access to the broad range of possible intermediates for C–O scission, C–C coupling and product release that are required to generate metal-free C<sub>2</sub>O<sub>1</sub> fragments at a single metal site, we used a ligand set with a propensity for diverse binding modes. The terphenyl–diphosphine ligand undergoes changes in the coordination mode of the central arene as a function of the oxidation state and primary coordination sphere of the metal<sup>24,25</sup>. Furthermore, the requirement for numerous reducing equivalents can be satisfied by arene reduction<sup>25</sup> in concert with oxidation-state changes at the metal. We anticipated that supporting molybdenum (Mo)—a metal known to enact both C–O cleavage<sup>21</sup> and C–C coupling<sup>17</sup>—on such a versatile ligand platform might provide access to the diverse intermediates necessary to perform mononuclear CO deoxygenation and coupling.

The Mo<sup>II</sup>–dicarbonyl complex (compound 1; reported in ref. 24) can be reduced in a stepwise fashion with potassium-intercalated graphite, KC<sub>8</sub> (or with potassium naphthalenide, K(C<sub>10</sub>H<sub>8</sub>)), to generate formal Mo<sup>0</sup> (2), Mo<sup>–II</sup> (3) and Mo<sup>–III</sup> (4) species (Fig. 1). These complexes, spanning six oxidation states, have all been fully characterized by single crystal X-ray diffraction (XRD). The change in ligand-binding mode in the solid state is notable (Fig. 2). Dication 1 displays the coordination of both phosphine arms, as well as an  $\eta^6$ –arene interaction; in compound 2, one phosphine is dissociated. Further reduction leads to the formation of potassium-ion-bridged polynuclear clusters, displaying interactions between the cations and the free phosphine arm, the carbonyl ligands, and the ligand  $\pi$ -systems. In these oligomers, the central arene is distorted considerably from planarity, and the molybdenum–arene binding mode changes from  $\eta^6$  in compound 2 to  $\eta^4$  in compound 3, and to  $\eta^3$  or  $\eta^4$  in the two unique subunits of 4. These differences, together with the specific localization of C–C single and double bond character, are consistent with substantial delocalization of reducing equivalents into the terphenyl scaffold. The ability of the pendant arene to act as a reservoir for electrons, adopting a partial cyclohexyldienylidide character<sup>26</sup>, is essential to supporting these uncommon highly reduced species<sup>27</sup>.

<sup>1</sup>Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 East California Boulevard, MC 127-72, Pasadena, California 91125, USA.





The overall reaction (shown at the top) involves the transformation of two

The overall reaction (shown at the top) involves the transformation of two Mo-bound carbonyls, with the addition of four reducing equivalents ( $e^-$ ) and four equivalents of electrophile ( $E^+$ ), to generate a metal-free  $C_2O_1$  product. A detailed scheme follows. Starting with compound **1**, successive electron loading (to **2**, then **3**, then **4**)—using  $KC_8$  and facilitated by electron storage in the pendant arene—leads to substantial CO activation. The addition of the silyl electrophile  $Me_3SiCl$  to **3** results in C–O cleavage and the formation of silylcarbyne **7**, proposed to proceed via a terminal molybdenum carbide, **8**. From **7**, two electrons are required for the formation of  $C_2$  products (**6b** and **6c**), which are spontaneously displaced by  $N_2$ , providing compound **5**. Addition of bulkier silyl electrophiles (*i*- $Pr_3SiCl$ ) to **3** or the more-reduced **4** results directly in the generation of compound **5** and a  $C_2$  organic fragment (**6a**). Synthesis of these  $C_2O_1$  products (**6a**, **6b** and **6c**) from two CO ligands represents an overall four-electron transformation.  $E^+$ , electrophile;  $e^-$ , electron; Mo, molybdenum; OTf, trifluoromethanesulfonate.

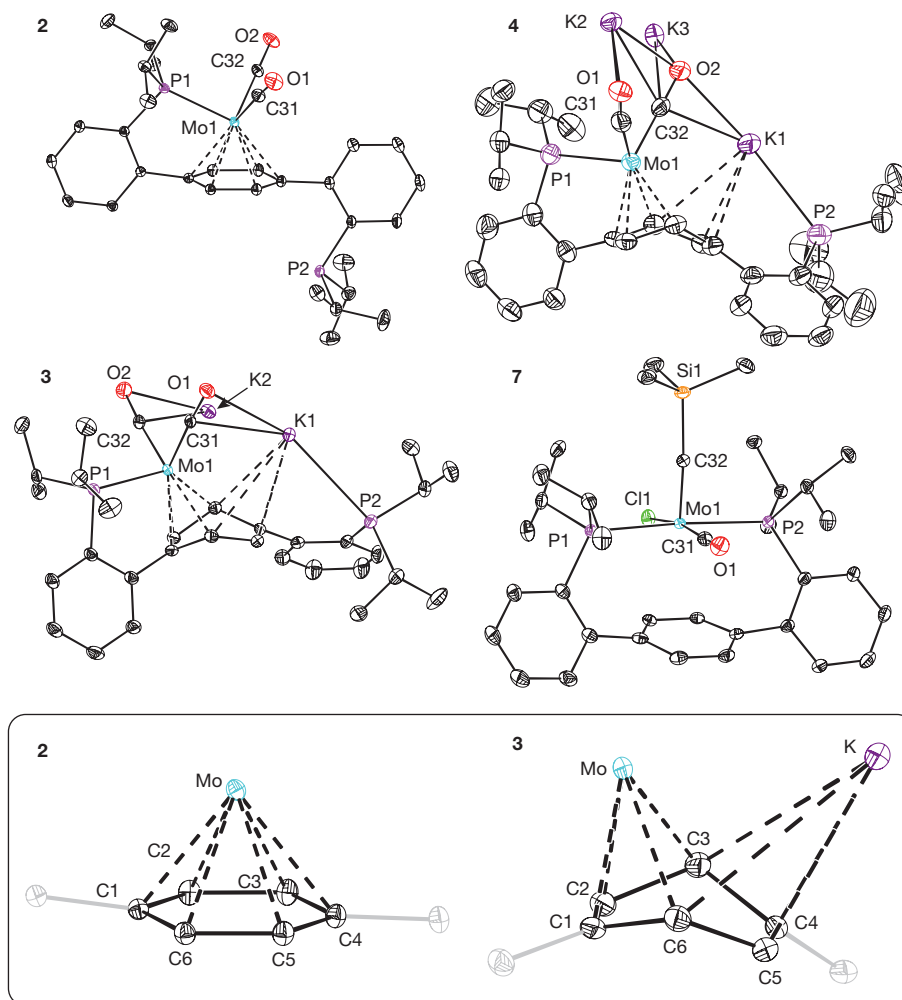
The elongation of C–O and contraction of Mo–CO bond lengths in the solid state (Supplementary Table 2) is consistent with appreciable back-bonding upon reduction; the delocalization of electron density into the carbonyl ligands likewise plays a pivotal role in stabilizing such reduced compounds. Attenuation of the CO stretching frequencies across the reduction series (for compound 1, 2,044 cm<sup>−1</sup> and 1,989 cm<sup>−1</sup>; 2, 1,887 cm<sup>−1</sup> and 1,832 cm<sup>−1</sup>;

3, 1,657 cm<sup>-1</sup> and 1,570 cm<sup>-1</sup>; **4**, less than 1,500 cm<sup>-1</sup>) further supports this claim; each set of stretches shifts as anticipated for the <sup>13</sup>C isotopologues. Given this strong CO activation, we explored electrophilic functionalization. Under an N<sub>2</sub> atmosphere, treating compound **1** with excess triisopropylsilyl chloride (*i*-Pr<sub>3</sub>SiCl) leads to the formation of N<sub>2</sub>-ligated complex **5** (ref. 24) as the major metal-containing product (75% yield, as determined by <sup>31</sup>P{<sup>1</sup>H} nuclear magnetic resonance (NMR) spectroscopy); partial oxidation, producing compound **2**, also occurs. Comparison with authentic samples confirmed hexa-*iso*-propyldisilyloxane ((*i*-Pr<sub>3</sub>Si)<sub>2</sub>O; by gas chromatography/mass spectrometry) and the C<sub>2</sub>O<sub>1</sub> fragment **6a** (by <sup>13</sup>C{<sup>1</sup>H} NMR spectroscopy) as the organic by-products generated in the transformation of **4** to compound **5** (Fig. 1); <sup>13</sup>C labelling confirmed that the C<sub>2</sub> unit originated from the carbonyl ligands. This remarkable reaction encompasses a trianionic metal complex that displays a high degree of CO activation, cleavage of the strong C–O bond, formation of a C≡C bond, and spontaneous product release. These features distinguish this system from reported examples<sup>7,8,11–17</sup> of transition-metal-mediated CO coupling, which lead to two electron-reduced products or prove resilient to dissociation of the CO cleavage fragments from the metal centre.

Balancing this reaction, however, reveals that an additional reducing equivalent is required, beyond the three that are stored in complex **4**. The generation of compound **2** suggests that trianion **4** may be acting as a sacrificial reductant. Consistent with this hypothesis, adding *i*-Pr<sub>3</sub>SiCl to the less-reduced compound **3** (Fig. 1) lowers the yield of coupling products (to 50%, as determined by <sup>31</sup>P{<sup>1</sup>H} NMR spectroscopy) and increases the oxidation that generates compound **2**. Performing the reaction with an additional electron equivalent—in the form of K(C<sub>10</sub>H<sub>8</sub>)—leads to greater than 90% conversion of **4** to compound **5** (as measured by NMR spectroscopy), with no observation of the oxidation product, **2**. Given this high yield, we targeted a closed synthetic cycle; however, addition of CO gas to **5** does not regenerate **2**.

Anticipating that a smaller silyl electrophile might facilitate the observation of reaction intermediates, we treated a thawing tetrahydrofuran solution of dianion **3** with excess trimethylsilyl chloride ( $\text{Me}_3\text{SiCl}$ ). The silylcarbyne complex **7** was formed cleanly (Fig. 1), as suggested by an alkylidyne resonance at 355.85 parts per million (p.p.m.) in the  $^{13}\text{C}\{^1\text{H}\}$  NMR spectrum, and confirmed by single-crystal XRD (Fig. 2). To support the higher oxidation state of molybdenum in **7**, the ligand adopts a pseudo-square pyramidal coordination environment, with no notable metal–arene interaction. Hexamethyldisiloxane ( $(\text{Me}_3\text{Si})_2\text{O}$ ) was detected in the reaction mixture by  $^{13}\text{C}\{^1\text{H}\}$  NMR spectroscopy, accounting for the cleaved oxygen atom. The formation of **7** from **3** represents a complete scission of the  $\text{C}\equiv\text{O}$  bond—a six-electron transformation. This conversion is accomplished in high yields, without additional reducing equivalents, owing to the ability of the supporting ligand to facilitate the storage of multiple electrons.

The formation of deoxygenated compound **7** suggests the intermediacy of a terminal molybdenum carbide complex, **8**, generated after loss of silyl ether. To probe the viability of this proposal, we targeted **8** by independent synthesis (Fig. 1). At low temperatures ( $-20^{\circ}\text{C}$ ), the desilylation of  $^{13}\text{C}$ -labelled **7** with tetra-*n*-butylammonium fluoride (*n*-Bu<sub>4</sub>NF) results in clean conversion to a new complex with a resonance in the  $^{13}\text{C}\{^1\text{H}\}$  NMR spectrum at 546.20 p.p.m. (Fig. 3). This chemical shift is diagnostic of terminal carbide species<sup>21,28–30</sup>, and indicates that such a moiety is accessible in our system, even though examples of neutral terminal transition metal carbides are exceedingly rare<sup>29,30</sup>, and unprecedented for molybdenum. The coupling pattern observed in the  $^{13}\text{C}\{^1\text{H}\}$  NMR spectrum is consistent with isotopically enriched carbide (546.20 p.p.m.) and CO (233.16 p.p.m.) ligands both being coordinated to the same metal centre. Although compound **8** decomposes to a complex mixture at room temperature, treatment with Me<sub>3</sub>SiCl at  $-80^{\circ}\text{C}$  leads to the clean formation



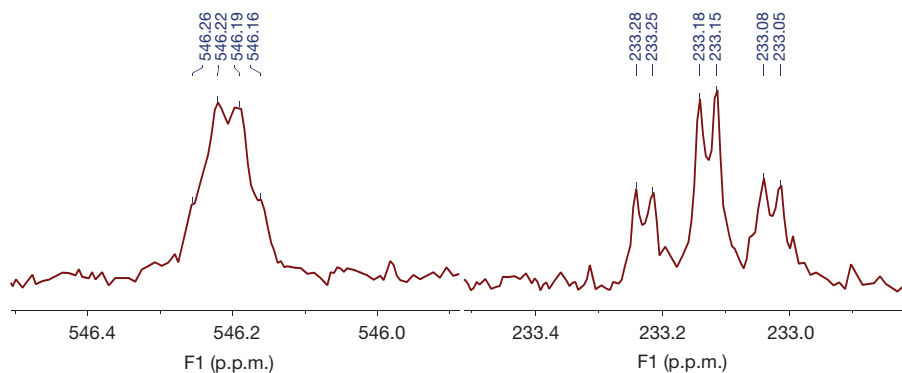
**Figure 2 | X-ray crystal structures of compounds 2, 3, 4 and 7.** Structures displaying a full molybdenum terphenyl-diphosphine unit are shown at the top; truncated enlargements of the molybdenum-arene cores of complexes 2 and 3 are shown in the inset. Reduction of compound 2 to generate 3 and 4 leads to deplanarization of the arene, consistent with a partial cyclohexyldienylidene character. In the more-oxidized

compound 7, no metal-arene interaction is observed. The molecular structures are displayed with anisotropic displacement ellipsoids shown at the 50% probability level. Co-crystallized solvent molecules, potassium-bound tetrahydrofuran molecules, and hydrogen atoms are omitted for clarity. A single molybdenum core is represented for the polynuclear clusters 3 and 4.

of 7 upon warming, showing that the carbide could be an intermediate en route to 7 from 3.

Recognizing that the silylcarbyne 7 incorporates both C-Si and C-O linkages, we explored the formation of C-C bonds from this species. Treating 7 with two equivalents of  $\text{KC}_8$  under  $\text{N}_2$  leads to the quantitative formation of compound 5 and an unbound  $\text{C}_2\text{O}_1$  fragment (6b);

Fig. 1). Subsequent addition of  $\text{Me}_3\text{SiCl}$  produces the bis(trimethylsilyl) ketene 6c. This reaction sequence demonstrates the ability of 7 to undergo reductive C-C coupling followed by dissociation of the partially deoxygenated  $\text{C}_2$  unit from molybdenum. An additional reducing equivalent is pre-loaded in the system when comparing trianion 4 to dianion 3, resulting directly in coupling chemistry from the former.



**Figure 3 | NMR spectroscopic data.** These  $^{13}\text{C}\{^1\text{H}\}$  NMR spectroscopic data (126 MHz, 25 °C) are for a solution of compound 8, bearing  $^{13}\text{C}$ -labels on the CO-derived carbon atoms, dissolved in tetrahydrofuran/benzene- $d_6$ . The coupling pattern ( $^2J(\text{C}, \text{C}) = 3.46 \text{ Hz}$ ;  $^2J(\text{P}, \text{C}) = 3.26 \text{ Hz}$ ;

$^2J(\text{P}, \text{CO}) = 12.55 \text{ Hz}$ ) and the chemical shifts of the isotopically enriched carbon atoms are consistent with the coordination of carbide (546.20 p.p.m.) and CO ligands (233.16 p.p.m.) to the same metal centre.

Our work establishes a series of design elements to be considered for monometallic complexes capable of the conversion of CO to C<sub>2</sub>O<sub>1</sub> products. Specifically, a supporting ligand that can store reducing equivalents in a pendant arene facilitates the activation and cleavage of the C–O bond, generating an intermediate molybdenum carbide. Subsequent C–C coupling is facilitated by silylation and reduction. This combination of bond breaking, bond forming, and organic-product release is unprecedented for a homogeneous monometallic system. Moreover, although metal carbides are commonly invoked intermediates for heterogeneous CO catenation, never before has a well-defined, terminal metal carbide complex been demonstrated as a CO-coupling intermediate. The electronically and coordinatively flexible ligand architecture, with variable arene-binding modes and a hemilabile phosphine donor, is instrumental in supporting these diverse moieties, facilitating the activation, scission, and coupling of CO to generate C<sub>2</sub>O<sub>1</sub> species. The transformations observed here provide precedent relevant to the conversion chemistry of oxygenated C<sub>1</sub> feedstocks to partially deoxygenated C<sub>2</sub> products, showcasing multi-electron reactions at a single metal site.

Received 17 July; accepted 6 October 2015.

Published online 21 December 2015.

- Appel, A. M. *et al.* Frontiers, opportunities, and challenges in biochemical and chemical catalysis of CO<sub>2</sub> fixation. *Chem. Rev.* **113**, 6621–6658 (2013).
- Costentin, C., Robert, M. & Saveant, J.-M. Catalysis of the electrochemical reduction of carbon dioxide. *Chem. Soc. Rev.* **42**, 2423–2436 (2013).
- Gattrell, M., Gupta, N. & Co, A. A review of the aqueous electrochemical reduction of CO<sub>2</sub> to hydrocarbons at copper. *J. Electroanal. Chem.* **594**, 1–19 (2006).
- Hu, Y. L., Lee, C. C. & Ribbe, M. W. Extending the carbon chain: hydrocarbon formation catalyzed by vanadium/molybdenum nitrogenases. *Science* **333**, 753–755 (2011).
- Li, C. W., Ciston, J. & Kanan, M. W. Electroreduction of carbon monoxide to liquid fuel on oxide-derived nanocrystalline copper. *Nature* **508**, 504–507 (2014).
- West, N. M., Miller, A. J. M., Labinger, J. A. & Bercaw, J. E. Homogeneous syngas conversion. *Coord. Chem. Rev.* **255**, 881–898 (2011).
- Gardner, B. M. *et al.* Homologation and functionalization of carbon monoxide by a recyclable uranium complex. *Proc. Natl Acad. Sci. USA* **109**, 9265–9270 (2012).
- Summerscales, O. T., Cloke, F. G. N., Hitchcock, P. B., Green, J. C. & Hazari, N. Reductive cyclotrimerization of carbon monoxide to the delatate dianion by an organometallic uranium complex. *Science* **311**, 829–831 (2006).
- Miller, R. L., Wolczanski, P. T. & Rheingold, A. L. Carbide formation via carbon monoxide dissociation across a tungsten-tungsten triple bond. *J. Am. Chem. Soc.* **115**, 10422–10423 (1993).
- LaPointe, R. E., Wolczanski, P. T. & Mitchell, J. F. Carbon monoxide cleavage by (silox)<sub>3</sub>Ta (silox = *tert*-Bu<sub>3</sub>SiO<sup>−</sup>). *J. Am. Chem. Soc.* **108**, 6382–6384 (1986).
- Evans, W. J., Grate, J. W., Hughes, L. A., Zhang, H. & Atwood, J. L. Reductive homologation of carbon monoxide to a ketenecarboxylate by a low-valent organolanthanide complex: synthesis and x-ray crystal structure of [(C<sub>5</sub>Me<sub>5</sub>)<sub>4</sub>Sm<sub>2</sub>(O<sub>2</sub>CCCO)(THF)]<sub>2</sub>. *J. Am. Chem. Soc.* **107**, 3728–3730 (1985).
- Ballmann, J., Pick, F., Castro, L., Fryzuk, M. D. & Maron, L. Cleavage of carbon monoxide promoted by a dinuclear tantalum tetrahydride complex. *Organometallics* **31**, 8516–8524 (2012).
- Watanabe, T., Ishida, Y., Matsuo, T. & Kawaguchi, H. Reductive coupling of six carbon monoxides by a ditantalum hydride complex. *J. Am. Chem. Soc.* **131**, 3474–3475 (2009).
- Shima, T. & Hou, Z. Hydrogenation of carbon monoxide by tetranuclear rare earth metal polyhydrido complexes. Selective formation of ethylene and isolation of well-defined polyoxo rare earth metal clusters. *J. Am. Chem. Soc.* **128**, 8124–8125 (2006).
- Matsuo, T. & Kawaguchi, H. A synthetic cycle for H<sub>2</sub>/CO activation and allene synthesis using recyclable zirconium complexes. *J. Am. Chem. Soc.* **127**, 17198–17199 (2005).
- Belmonte, P. A., Cloke, F. G. N. & Schrock, R. R. Reduction of carbon monoxide by binuclear tantalum hydride complexes. *J. Am. Chem. Soc.* **105**, 2643–2650 (1983).
- Carnahan, E. M., Protasiewicz, J. D. & Lippard, S. J. 15 years of reductive coupling—what have we learned? *Acc. Chem. Res.* **26**, 90–97 (1993).
- Büchner, W. & Weiss, E. Zur kenntnis sogenanntes alkalicarbonyl 4 über reaktion von geschmolzenem kalium mit kohlenmonoxid. *Helv. Chim. Acta* **47**, 1415–1423 (1964).
- Wayland, B. B., Sherry, A. E. & Coffin, V. L. Selective reductive coupling of carbon-monoxide. *J. Chem. Soc. Chem. Commun.* 662–663 (1989).
- Suess, D. L. M. & Peters, J. C. A CO-derived iron dicarbonyl that releases olefin upon hydrogenation. *J. Am. Chem. Soc.* **135**, 12580–12583 (2013).
- Peters, J. C., Odom, A. L. & Cummins, C. C. A terminal molybdenum carbide prepared by methylidyne deprotonation. *Chem. Commun.* 1995–1996 (1997).
- Kreissl, F. R., Frank, A., Schubert, U., Lindner, T. L. & Huttner, G. Carbonyl-η-cyclopentadienyl-(4-methylphenylketenyl)-bis(trimethylphosphane) tungsten—a novel, stable transition metal-substituted ketene. *Angew. Chem. Int. Edn* **15**, 632–633 (1976).
- Churchill, M. R., Wasserman, H. J., Holmes, S. J. & Schrock, R. R. Coupling of methylidyne and carbonyl ligands on tungsten. Crystal structure of W(η<sup>2</sup>-HC≡COAlCl<sub>3</sub>)(CO)(PMe<sub>3</sub>)<sub>3</sub>Cl. *Organometallics* **1**, 766–768 (1982).
- Buss, J. A., Edouard, G. A., Cheng, C., Shi, J. & Agapie, T. Molybdenum catalyzed ammonia borane dehydrogenation: oxidation state specific mechanisms. *J. Am. Chem. Soc.* **136**, 11272–11275 (2014).
- Horak, K. T., Velian, A., Day, M. W. & Agapie, T. Arene non-innocence in dinuclear complexes of Fe, Co, and Ni supported by a *para*-terphenyl diphosphine. *Chem. Commun.* **50**, 4427–4429 (2014).
- Cassani, M. C., Gun'ko, Y. K., Hitchcock, P. B., Lappert, M. F. & Laschi, F. Synthesis and characterization of organolanthanidocene(III) (Ln = La, Ce, Pr, Nd) complexes containing the 1,4-cyclohexa-2,5-dienyl ligand (benzene 1,4-dianion): structures of [K([18]-crown-6)][Ln(η<sup>3</sup>-C<sub>5</sub>H<sub>3</sub>(SiMe<sub>3</sub>)<sub>2</sub>-1,3)<sub>2</sub>(C<sub>6</sub>H<sub>6</sub>)] [Cp' = η<sup>3</sup>-C<sub>5</sub>H<sub>3</sub>(SiMe<sub>3</sub>)<sub>2</sub>-1,3; Ln = La, Ce, Nd]. *Organometallics* **18**, 5539–5547 (1999).
- Ellis, J. E. Adventures with substances containing metals in negative oxidation states. *Inorg. Chem.* **45**, 3167–3186 (2006).
- Enriquez, A. E., White, P. S. & Templeton, J. L. Reactions of an amphoteric terminal tungsten methylidyne complex. *J. Am. Chem. Soc.* **123**, 4992–5002 (2001).
- Carlson, R. G. *et al.* The metathesis-facilitated synthesis of terminal ruthenium carbide complexes: a unique carbon atom transfer reaction. *J. Am. Chem. Soc.* **124**, 1580–1581 (2002).
- Stewart, M. H., Johnson, M. J. A. & Kampf, J. W. Terminal carbido complexes of osmium: synthesis, structure, and reactivity comparison to the ruthenium analogues. *Organometallics* **26**, 5102–5110 (2007).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** We thank L. M. Henling and M. K. Takase for crystallographic assistance and D. VanderVelde for NMR expertise. We are grateful to Caltech and the National Science Foundation (grant CHE-1151918 to T.A., and GRFP to J.A.B.) for funding.

**Author Contributions** J.A.B. and T.A. designed the research. J.A.B. conducted the experiments. J.A.B. and T.A. interpreted the data and wrote the manuscript.

**Author Information** X-ray crystallographic coordinates for compounds **2**, **3**, **4** and **7** have been deposited at the Cambridge Crystallographic Database under accession numbers 1412068, 1412062, 1412063 and 1412064 respectively. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.A. ([agapie@caltech.edu](mailto:agapie@caltech.edu)).



# Slab melting as a barrier to deep carbon subduction

Andrew R. Thomson<sup>1,2</sup>, Michael J. Walter<sup>1</sup>, Simon C. Kohn<sup>1</sup> & Richard A. Brooker<sup>1</sup>

**Interactions between crustal and mantle reservoirs dominate the surface inventory of volatile elements over geological time, moderating atmospheric composition and maintaining a life-supporting planet<sup>1</sup>. While volcanoes expel volatile components into surface reservoirs, subduction of oceanic crust is responsible for replenishment of mantle reservoirs<sup>2,3</sup>. Many natural, 'superdeep' diamonds originating in the deep upper mantle and transition zone host mineral inclusions, indicating an affinity to subducted oceanic crust<sup>4–7</sup>. Here we show that the majority of slab geotherms will intersect a deep depression along the melting curve of carbonated oceanic crust at depths of approximately 300 to 700 kilometres, creating a barrier to direct carbonate recycling into the deep mantle. Low-degree partial melts are alkaline carbonatites that are highly reactive with reduced ambient mantle, producing diamond. Many inclusions in superdeep diamonds are best explained by carbonate melt–peridotite reaction. A deep carbon barrier may dominate the recycling of carbon in the mantle and contribute to chemical and isotopic heterogeneity of the mantle reservoir.**

Altered oceanic crust incorporates appreciable carbon, which is added by magmatic and hydrothermal processes<sup>8</sup>, and by addition of CO<sub>2</sub> during interaction of basalt with sea water<sup>9</sup>. Together, these alteration processes result in subducting lithosphere that contains an average of ~2 weight per cent (wt%) CO<sub>2</sub> in the uppermost volcanic section and 100–5,000 p.p.m. CO<sub>2</sub> throughout the remaining 7 km of crust<sup>8</sup>. Crustal carbon initially contains a mixture of reduced hydrocarbons<sup>8</sup> and oxidized carbonates<sup>9</sup>. However, metamorphic re-equilibration of slab carbon with ferric iron and/or oxidizing fluids produced during serpentine dehydration at sub-arc conditions probably converts most slab carbon to carbonate<sup>10</sup>. Some of this carbon is returned to the exosphere in volcanic arcs, but both theoretical<sup>11</sup> and experimental<sup>12</sup> studies suggest that a considerable quantity of carbon may survive beyond slab dehydration, and be subducted into the mantle.

Carbon is insoluble in mantle silicate minerals<sup>13</sup> and is stored either as carbonate, carbide or diamond, depending on the oxidation state. Under oxidizing conditions, carbonate lowers the melting point (solidus) of mantle peridotite by up to 500 °C compared with volatile-free mantle<sup>14</sup>. However, at the more reducing conditions prevailing deeper in the upper mantle and transition zone, carbon will be stored as diamond or carbide minerals<sup>15</sup>, where it does not appreciably influence melting.

Superdeep diamonds originate from depths beneath the lithospheric mantle (>200 km) and are the only direct samples of the deep mantle carbon reservoir. Inclusions in these diamonds are dominated by upper mantle and transition zone minerals, which are mostly associated with subducted mafic lithologies rather than peridotite<sup>4–7,16</sup>. Many superdeep diamonds are made of isotopically light carbon<sup>6,7</sup> and, where measured, their inclusions contain isotopically heavy oxygen<sup>17</sup>, unambiguously indicating an origin from recycled surface material<sup>6,7,17</sup>. The elevated trace element abundances of many silicate inclusions suggest crystallization from a low-degree melt, thought to be generated from melting of subducted oceanic crust<sup>7,18</sup>. Here we examine the fate of subducting carbonated mid-ocean-ridge basalt (MORB) as it reaches the transition zone, and the potential for

melt–mantle reactions to reproduce superdeep diamonds and their distinctive inclusion assemblages.

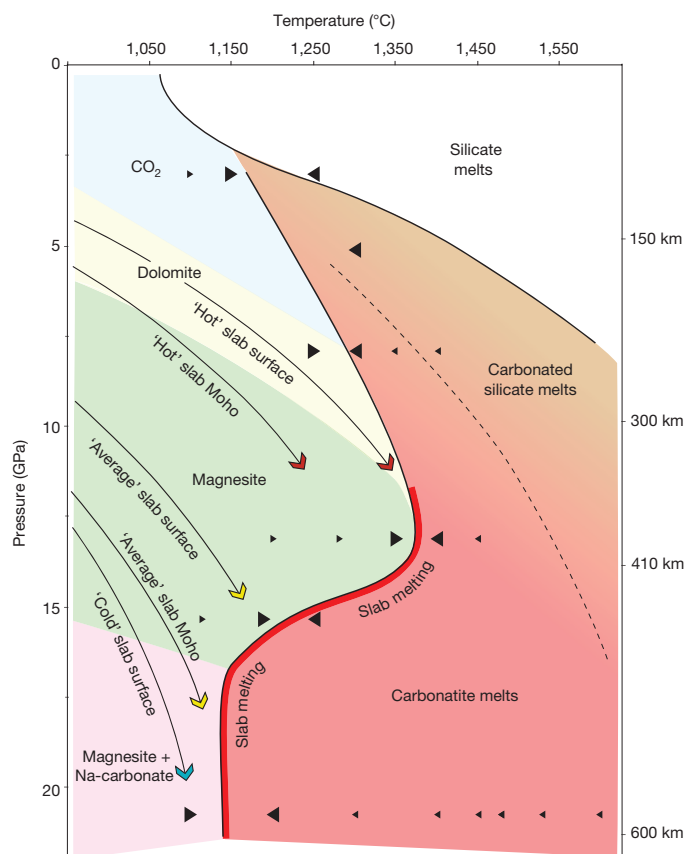
Previous experimental studies have investigated the melting behaviour of carbonated basalt at elevated pressures, but only one extends beyond 10 GPa (ref. 19). These studies show a remarkable diversity in melting behaviour, making extrapolation to higher pressures difficult. In addition, the bulk compositions employed in previous studies often contain considerably more CO<sub>2</sub> than mean oceanic crust, and fall outside the compositional field of natural MORB rocks (see Methods, Extended Data Fig. 1 and Extended Data Table 1). To understand better the melting behaviour of deeply subducted oceanic crust, we determined the melting phase relations of a synthetic MORB composition containing 2.5 wt% CO<sub>2</sub> between 3 and 21 GPa (Methods). Our starting composition replicates the major element composition of basaltic rocks from International Ocean Discovery Program (IODP) hole 1256D<sup>20</sup> and falls within the range of natural crust compositions<sup>21</sup> (Extended Data Fig. 1).

We observe subsolidus phase assemblages containing garnet, clinopyroxene, a SiO<sub>2</sub> polymorph, and Ti-rich oxide at all pressures (Extended Data Figs 2, 3 and Extended Data Table 2). The carbon component was either CO<sub>2</sub>, dolomite, magnesite or magnesite plus Na-carbonate depending on the pressure, and the positions of solid carbonate phase boundaries are consistent with previous studies<sup>22,23</sup>. Near-solidus partial melts are CO<sub>2</sub>-bearing silicate melts below 7 GPa, and silica-poor calcic carbonatites above 7 GPa. The alkali component of carbonatite melts increases with pressure (Extended Data Fig. 4), and all melts have high TiO<sub>2</sub>/SiO<sub>2</sub> (see Methods and Extended Data Figs 2–5, Extended Data Table 2 and Supplementary Tables 1–4 for detailed results).

The melting temperature of carbonated oceanic crust is tightly bracketed from ~3 to 21 GPa (Fig. 1). Melting temperatures increase steadily with increasing pressure until about 13 GPa, when the solidus dramatically drops over a narrow pressure interval by ~200 °C. This drop in solidus temperature is caused by a change in clinopyroxene composition towards a more sodium-rich composition above 13 GPa due to dissolution of sodium-poor pyroxene components into coexisting garnet. Eventually, clinopyroxene becomes so sodium-rich that a coexisting Na-carbonate mineral ([Na<sub>0.97</sub>K<sub>0.03</sub>]<sub>0.33</sub>[Ca<sub>0.86</sub>Mg<sub>0.11</sub>Fe<sub>0.03</sub>]<sub>0.67</sub>CO<sub>3</sub>) stabilizes in the subsolidus assemblage, causing the depression along the solidus. The loss of the sodium-poor clinopyroxene component, and the extended stability of sodic clinopyroxene in the absence of an alternative sodium-bearing silicate phase, is consistent with previous studies<sup>24</sup>. Above 16 GPa the solidus changes little with pressure, remaining at ~1,150 °C, consistent with the solidus observed in a sodium-rich simplified system in which sodic carbonate ([Na,K]<sub>0.33</sub>Ca<sub>0.67</sub>CO<sub>3</sub>) controls melting temperatures<sup>25</sup>. The major difference between this work and the previous study of carbonated MORB above 8 GPa (ref. 19) is the different phase assemblage resulting from the lower and more realistic CO<sub>2</sub> and CaO contents of our bulk composition. Previous bulk compositions with higher CaO contents (Extended Data Figs 1 and 5) are located on the calcium-rich side of the majorite–clinopyroxene tie-line and stabilize aragonite as the carbon-hosting phase, which can incorporate

<sup>1</sup>School of Earth Sciences, University of Bristol, Bristol BS8 1RJ, UK. <sup>2</sup>Department of Earth Sciences, University College London, London WC1E 6BT, UK.





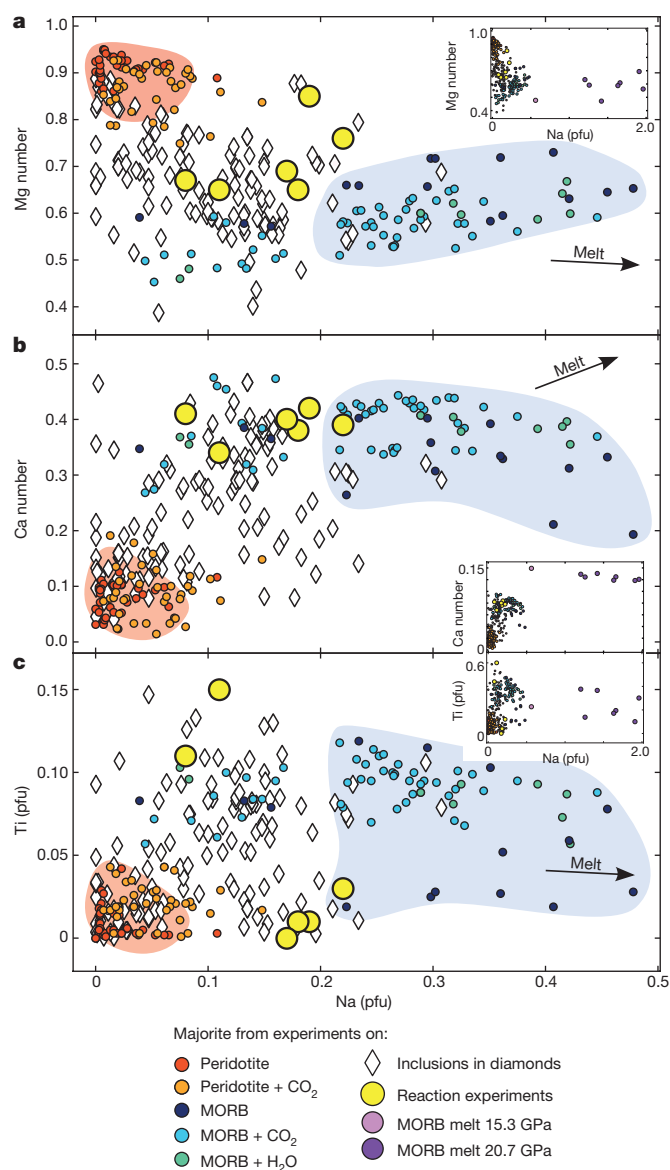
**Figure 1 | The melting curve of carbonated MORB compared to hot and cold subduction geotherms<sup>26</sup>.** The stability fields of carbon-bearing phases are identified in different colours. Experiments performed are marked by filled triangles indicating their relationship to the solidus, larger symbols mark solidus brackets. The solidus ledge creates a narrow depth interval where slab temperatures intersect the melting curve, producing a focused region of melt generation at the top of the transition zone.

considerable Na<sub>2</sub>O. The lower CO<sub>2</sub> content in our bulk composition results in a smaller proportion of carbonate, of which the dominant species is sodium-poor magnesite. Thus, sodic clinopyroxene remains stable as an alkali-host, coexisting with stoichiometric Na-carbonate to high pressures.

The deep solidus depression in carbonated oceanic crust at uppermost transition zone conditions creates a key control on the recycling of mantle carbon. Extrapolation of the range of modern-day oceanic crustal geotherms into the transition zone<sup>26</sup> reveals that the majority of slabs will intersect our solidus for carbonated recycled MORB (Fig. 1), producing carbonatite melt. Given the expected temperature profile in the average subducted slab<sup>26</sup> we estimate that melting would occur to depths of at least 7 km into the crustal section. Only the coldest modern-day slabs escape the solidus depression and are able to carry their carbonate cargo beyond the transition zone. If ancient slabs were hotter<sup>3</sup>, it seems likely that carbonate subduction through the transition zone and into the lower mantle has been limited throughout Earth's history. While the natural variability of subducting slabs (for example, composition, age, temperature) will have created some range in melting behaviour, the depression of the carbonated eclogite solidus will remain an efficient barrier. Thus, direct recycling of carbon into the lower mantle may have been highly restricted throughout most of the Earth's history, instead being redistributed throughout the upper mantle.

Carbonatitic melts are predicted to be mobile at mantle conditions due to their low viscosity and ability to wet silicate minerals<sup>27</sup>, so should percolate out of the slab and infiltrate the overlying peridotitic mantle<sup>25</sup>. Experiments suggest that below ~250 km, ambient

mantle oxygen fugacity is reducing, and a free metal phase may be present in the mantle<sup>28</sup>. Under such conditions a carbonate melt is unstable and will reduce to diamond plus oxygen by a 'redox-freezing' reaction<sup>28</sup> such as:  $\text{MgCO}_3 + 2\text{Fe}^0 = 3(\text{Mg}_{0.33}, \text{Fe}_{0.67})\text{O} + \text{C}$ . Thus, the expulsion of carbonatite melts due to melting of oceanic crust along the solidus depression provides an ideal environment for diamond growth across a depth interval of ~300–700 km. We predict that the interaction between MORB-derived carbonatite melt and ambient peridotite is capable of reproducing many of the characteristics of superdeep diamonds and the mineral inclusions that they capture from this depth interval<sup>4,5</sup>. The most common silicate minerals identified in superdeep diamonds are majorite garnet, and a titanium-bearing, calcium-silicate phase commonly interpreted as retrogressed 'calcium perovskite'<sup>4,6,7,18</sup>. Barometric estimates of the crystallization pressures for these majorite inclusions indicate they crystallized between 10 and

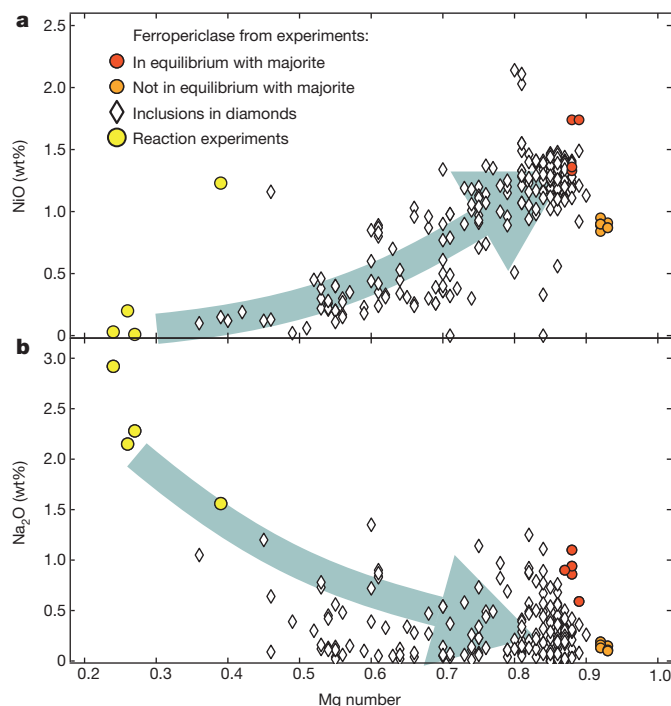


**Figure 2 | Composition of majoritic garnet minerals from previous experimental studies, inclusions in diamonds and reaction experiments.** a–c, The red field outlines the approximate range of peridotitic majorite compositions, the blue field outlines the range of MORB majorites from pressures above the carbonated MORB solidus ledge ( $\geq 9$  GPa). Na (per formula unit (pfu)) plotted against Mg number (Mg number =  $\text{Mg}/[\text{Mg} + \text{Fe}]$ ) (a), Ca number (Ca number =  $\text{Ca}/[\text{Ca} + \text{Mg} + \text{Fe}]$ ) (b), and Ti (pfu) (c). Data and corresponding references for this figure are provided in the online source data file.

16 GPa (ref. 5), and inclusions of calcium perovskite are constrained by their chemistry to have formed between  $\sim 10$  and 20 GPa (refs 6, 18). These pressures are remarkably consistent with the range of pressures at which slab crustal geotherms are predicted to intersect the carbonated solidus depression (Fig. 1).

Redox reactions in the mantle are complex and involve silicates, many containing iron that exists in both ferrous ( $\text{Fe}^{2+}$ ) and ferric form ( $\text{Fe}^{3+}$ ). To test the melt–mantle interaction model, we recreated the infiltration process in a second set of experiments by partially equilibrating a model slab melt with an iron-metal-bearing transition zone peridotite assemblage at 20 GPa (see Methods for details). We observe a reaction zone between the alkaline carbonatite melt and the initial peridotitic assemblage of majorite, wadsleyite, calcium-silicate perovskite and iron metal that consists of sodium-rich majoritic garnet,  $\text{Ca}[\text{Si,Ti}]\text{O}_3$  perovskite, ferrous ringwoodite (Mg number  $\sim 75$ ), ferropericlasite (Mg number  $\sim 0.4$ ) and diamond (Extended Data Figs 6, 7 and Extended Data Table 3). We compare the resulting mineral compositions with previous experimental data for peridotite and MORB systems to investigate whether natural inclusion assemblages might preserve a record of mineral–melt reactions.

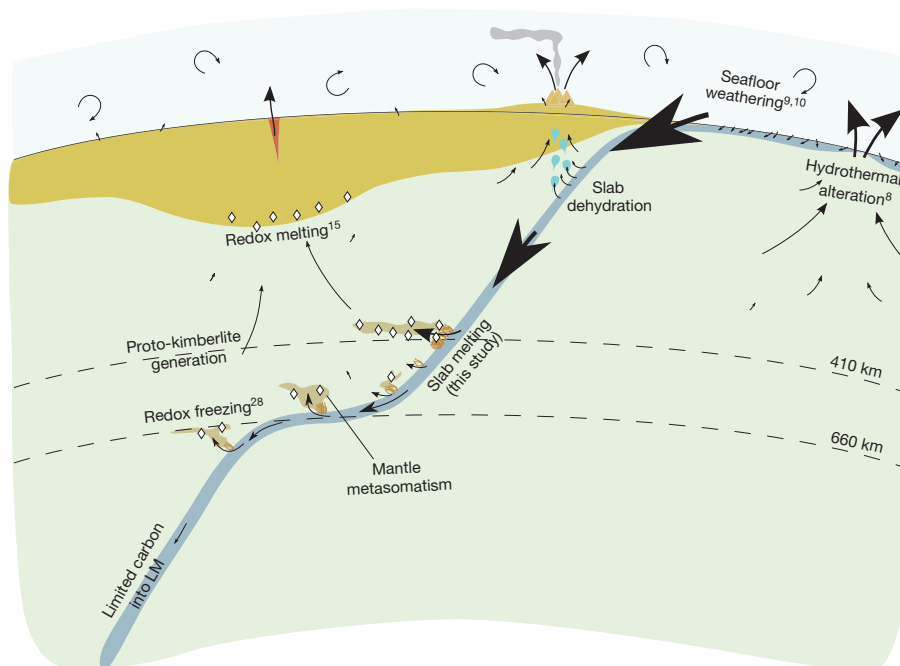
The compositions of the majority of superdeep majoritic garnet inclusions are not typical of those expected in either peridotitic or eclogitic bulk compositions (Fig. 2) and instead lie between these two end-members. These intermediate compositions have previously been described as pyroxenitic, and it was suggested that the transition zone may harbour a large component of this rock type<sup>16</sup>. Our results suggest an alternative explanation. In Fig. 2 the majoritic garnets produced during the experimental melt–mantle interaction are intermediate between peridotitic and eclogitic compositions, and cover much of the range seen in the diamond inclusions. The chemical imprint imparted by the MORB carbonatite on the peridotitic mantle is recorded in the inclusions as elevated Ca number, Na and Ti contents alongside depleted Mg number. Our experiments only demonstrate the composition of garnets produced near the beginning of the melt–mantle interaction sequence, and we suggest that the intermediate character of the natural inclusions records a snapshot



**Figure 3 | Composition of ferropericlasite minerals from previous experimental studies, inclusions in diamonds and reaction experiments.** a, b, Blue arrows indicate the compositional evolution expected as melt–mantle interactions progress. Data and corresponding references are provided in the online source data file.

of the infiltration and reaction of slab-derived carbonatite melt with peridotite.

Experimental calcium perovskites have high titanium ( $\sim 40$ – $60$  mol%  $\text{CaTiO}_3$ ) and are essentially magnesium-free, features observed throughout the global range of ‘calcium perovskite’ inclusions



**Figure 4 | Schematic of the deep mantle carbon cycle.** Arrows represent paths and estimates of the relative magnitudes of carbon fluxes. Downwelling slabs dehydrate at sub-arc depths but retain the majority of their carbon cargo. Upon reaching the transition zone they produce carbonatite melts (this study) along the solidus ledge that

infiltrate<sup>28</sup> and react with the overlying mantle (this study). This causes diamond production, refertilization and associated metasomatism of the surrounding mantle. The melting of recycled crust in the transition zone essentially prevents carbon transport into the lower mantle (LM).

(Extended Data Fig. 8). Thus, our reaction experiments reproduce the unique characteristics of diamond-hosted 'calcium perovskite' inclusions. Crystallization by reaction between a low-degree carbonated melt and peridotite is also consistent with the extremely elevated trace element contents of diamond-hosted 'calcium perovskites' inclusions<sup>24</sup>.

Probably the most abundant inclusions in superdeep diamonds are magnesium-iron oxide ([Mg,Fe]O), which are often interpreted to indicate diamond growth in the lower mantle<sup>4</sup>. However, our experiments demonstrate that ferropericlase can be produced in reactions between carbonatitic melt and reduced mantle peridotite at upper mantle pressures rather than requiring a lower mantle origin<sup>29</sup>. Figure 3 demonstrates that natural ferropericlase inclusions are almost all iron-rich relative to ferropericlase expected in mantle peridotite, and their compositions form arrays towards higher NiO and lower Na<sub>2</sub>O with increasing magnesium number. Our experimental ferropericlase compositions lie at the end of the arrays and are iron-rich because the peridotite starting material was initially iron-saturated. We suggest that, like the majorite inclusions, the array of intermediate ferropericlase compositions record the progressive reaction of carbonatite melt and ambient mantle.

The melting-phase relations of recycled oceanic crust suggest that slabs should undergo melting and loss of carbonate components in the transition zone (Fig. 4), a process that has considerable implications for the deep carbon cycle. The compositions of diamond-hosted inclusions provide strong evidence of this process and confirm that carbon must survive subduction beyond sub-arc dehydration reactions. We predict that carbon is rarely transported beyond the transition zone and instead refertilizes the upper mantle as diamond. Oxidation of diamond-bearing mantle upon upwelling can lead to redox melting<sup>15</sup> beneath the lithosphere and contribute markedly to the generation and geochemical signature of surface lavas. This process also probably contributes to the formation of distinctive chemical and isotopic reservoirs in the mantle<sup>30</sup>. Superdeep diamonds provide a physical record of carbon recycling above subducting slabs, which can be used to infer the residence time of carbon in the mantle. This residence time is regulated by rates of subduction, convective mantle upwelling and melting beneath the lithosphere, and could occur over a range of timescales, perhaps as short as tens to hundreds of millions of years, suggesting the mantle carbon cycle can be considerably more vigorous than previously estimated<sup>2,3</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 3 March; accepted 19 October 2015.**

- Zahnle, K. *et al.* Emergence of a habitable planet. *Space Sci. Rev.* **129**, 35–78 (2007).
- Sleep, N. H. & Zahnle, K. Carbon dioxide cycling and implications for climate on ancient Earth. *J. Geophys. Res.* **106**, 1373–1399 (2001).
- Dasgupta, R. & Hirschmann, M. M. The deep carbon cycle and melting in Earth's interior. *Earth Planet. Sci. Lett.* **298**, 1–13 (2010).
- Harte, B. Diamond formation in the deep mantle: the record of mineral inclusions and their distribution in relation to mantle dehydration zones. *Mineral. Mag.* **74**, 189–215 (2010).
- Stachel, T. Diamonds from the asthenosphere and the transition zone. *Eur. J. Mineral.* **13**, 883–892 (2001).
- Thomson, A. R. *et al.* Origin of sub-lithospheric diamonds from the Juina-5 kimberlite (Brazil): constraints from carbon isotopes and inclusion compositions. *Contrib. Mineral. Petrol.* **168**, 1081 (2014).
- Bulanova, G. *et al.* Mineral inclusions in sublithospheric diamonds from Collier 4 kimberlite pipe, Juina, Brazil: subducted protoliths, carbonated melts and primary kimberlite magmatism. *Contrib. Mineral. Petrol.* **160**, 489–510 (2010).
- Shilobreeva, S., Martinez, I., Busigny, V., Agrinier, P. & Laverne, C. Insights into C and H storage in the altered oceanic crust: results from ODP/IODP Hole 1256D. *Geochim. Cosmochim. Acta* **75**, 2237–2255 (2011).
- Alt, J. & Teagle, D. The uptake of carbon during alteration of ocean crust. *Geochim. Cosmochim. Acta* **63**, 1527–1535 (1999).

- Debret, B. *et al.* Redox state of iron during high-pressure serpentinite dehydration. *Contrib. Mineral. Petrol.* **169**, 36 (2015).
- Kerrick, D. M. & Connolly, J. A. D. Metamorphic devolatilization of subducted mid-ocean ridge metabasalts: implications for seismicity, arc magmatism and volatile recycling. *Earth Planet. Sci. Lett.* **189**, 19–29 (2001).
- Poli, S., Franzolin, E., Fumagalli, P. & Crottini, A. The transport of carbon and hydrogen in subducted oceanic crust: an experimental study to 5 GPa. *Earth Planet. Sci. Lett.* **278**, 350–360 (2009).
- Shcheka, S. S., Wiedenbeck, M., Frost, D. J. & Keppler, H. Carbon solubility in mantle minerals. *Earth Planet. Sci. Lett.* **245**, 730–742 (2006).
- Ghosh, S., Ohtani, E., Litasov, K. & Terasaki, H. Solidus of carbonated peridotite from 10 to 20 GPa and origin of magnesio-carbonatite melt in the Earth's deep mantle. *Chem. Geol.* **262**, 17–28 (2009).
- Stagno, V., Ojwang, D. O., McCammon, C. A. & Frost, D. J. The oxidation state of the mantle and the extraction of carbon from Earth's interior. *Nature* **493**, 84–88 (2013).
- Kiseeva, E. S. *et al.* Metaproxenite in the mantle transition zone revealed from majorite inclusions in diamonds. *Geology* **41**, 883–886 (2013).
- Ickert, R. B., Stachel, T., Stern, R. A. & Harris, J. W. Extreme <sup>18</sup>O-enrichment in majorite constrains a crustal origin of transition zone diamonds. *Geochim. Persp. Lett.* **1**, 65–74 (2015).
- Walter, M. J. *et al.* Primary carbonatite melt from deeply subducted oceanic crust. *Nature* **454**, 622–625 (2008).
- Kiseeva, E. S., Litasov, K. D., Yaxley, G. M., Ohtani, E. & Kamenetsky, V. S. Melting and phase relations of carbonated eclogite at 9–21 GPa and the petrogenesis of alkali-rich melts in the deep mantle. *J. Petrol.* **54**, 1555–1583 (2013).
- Expedition 309/312 Scientists. Site 1256. In *Superfast Spreading Rate Crust 2 and 3* (eds Teagle, D. A. H. *et al.*) Vol. 309/312 of *Proc. IODP* <http://dx.doi.org/10.2204/iodp.proc.309312.103.2006> (Integrated Ocean Drilling Program Management International, Inc., 2006).
- Gale, A., Dalton, C. A., Langmuir, C. H., Su, Y. & Schilling, J.-G. The mean composition of ocean ridge basalts. *Geochim. Geophys. Geosyst.* **14**, 489–518 (2013).
- Martin, A. M., Laporte, D., Koga, K. T., Kawamoto, T. & Hammouda, T. Experimental study of the stability of a dolomite + coesite assemblage in contact with peridotite: implications for sediment-mantle interaction and diamond formation during subduction. *J. Petrol.* **53**, 391–417 (2012).
- Luth, R. W. Experimental determination of the reaction aragonite + magnesite = dolomite at 5 to 9 GPa. *Contrib. Mineral. Petrol.* **141**, 222–232 (2001).
- Okamoto, K. & Maruyama, S. The Eclogite–Garnetite transformation in the MORB + H<sub>2</sub>O system. *Phys. Earth Planet. Inter.* **146**, 283–296 (2004).
- Litasov, K., Shatskiy, A., Ohtani, E. & Yaxley, G. Solidus of alkaline carbonatite in the deep mantle. *Geology* **41**, 79–82 (2013).
- Syracuse, E. M., van Keken, P. E. & Abers, G. A. The global range of subduction zone thermal models. *Phys. Earth Planet. Inter.* **183**, 73–90 (2010).
- Hammouda, T. & Laporte, D. Ultrafast mantle impregnation by carbonatite melts. *Geology* **28**, 283–285 (2000).
- Rohrbach, A. & Schmidt, M. W. Redox freezing and melting in the Earth's deep mantle resulting from carbon–iron redox coupling. *Nature* **472**, 209–212 (2011).
- Brey, G. P., Bulatov, V., Gurnis, A., Harris, J. W. & Stachel, T. Ferropericlase—a lower mantle phase in the upper mantle. *Lithos* **77**, 655–663 (2004).
- Jackson, M. G. & Dasgupta, R. Compositions of HIMU, EM1 and EM2 from global trends between radiogenic isotopes and major elements in ocean island basalts. *Earth Planet. Sci. Lett.* **276**, 175–186 (2008).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** A.R.T. acknowledges the support of NERC grant NE/J500033/1. M.J.W. and S.C.K. acknowledge the support of NERC grant NE/J008583/1. We thank S. Kearns and B. Buse for their assistance performing electron probe microanalyses and J. Blundy for contributing ideas and expertise during discussions with the authors.

**Author Contributions** A.R.T. designed, performed and analysed the experiments, gathered data from the literature and wrote the manuscript as part of his PhD studies. M.J.W. and S.C.K. provided training in experimental techniques, assisted during interpretation of results, provided advice and assisted with manuscript preparation in their roles as A.R.T.'s PhD supervisors. R.A.B. provided training and assistance with experimental techniques and sample preparation alongside contributing to the scientific content and preparation of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.R.T. ([a.r.thomson@ucl.ac.uk](mailto:a.r.thomson@ucl.ac.uk)).



## METHODS

**Starting materials.** The starting material for experiments to determine the melting-phase relations of carbonated MORB (ATCM1) replicates basalts from the IODP 1256D from the Eastern Pacific Rise<sup>20</sup> (the reported composition of IODP 1256D basalts is the average of all analyses presented in table T17 of ref. 20) with an added 2.5 wt% CO<sub>2</sub> (Extended Data Table 1). This material was formed by mixing high-purity SiO<sub>2</sub>, TiO<sub>2</sub>, Al<sub>2</sub>O<sub>3</sub>, FeO, MnO, MgO, Ca<sub>3</sub>(PO<sub>4</sub>)<sub>2</sub> and CaCO<sub>3</sub>, which were fired overnight at temperatures of 400–1,000 °C, of appropriate weights in an agate mortar under ethanol. This mixture was decarbonated and fused into a crystal-free glass in a one-atmosphere tube furnace by incrementally increasing the temperature from 400 to 1,500 °C before drop quenching into water. Subsequently weighed amounts of CaCO<sub>3</sub>, Na<sub>2</sub>CO<sub>3</sub> and K<sub>2</sub>CO<sub>3</sub> were ground into the glass, introducing the alkali and CO<sub>2</sub> components. After creation, the starting material was stored at 120 °C to avoid absorption of atmospheric water. Starting material ATCM2 replicates the near-solidus melt composition measured in melting experiments at 20.7 GPa and 1,400/1,480 °C. This was created by grinding natural magnesite and synthetic siderite with high-purity CaCO<sub>3</sub>, Na<sub>2</sub>CO<sub>3</sub>, K<sub>2</sub>CO<sub>3</sub>, SiO<sub>2</sub>, TiO<sub>2</sub>, Al<sub>2</sub>O<sub>3</sub> and Ca<sub>3</sub>(PO<sub>4</sub>)<sub>2</sub>. Synthetic siderite was created in a cold-seal pressure vessel experiment run at 2 kbar and 375 °C for 7 days. A double Au capsule design containing iron (II) oxalate dehydrate in the inner and a 1:1 mixture of CaCO<sub>3</sub> and SiO<sub>2</sub> in the outer capsule produced a pale beige powder confirmed as siderite using Raman spectroscopy. The material for a sandwich experiment, to ensure near-solidus melt compositions were accurately determined at 20.7 GPa, was formed of a 3:1 mixture of ATCM1:ATCM2.

The transition-zone peridotite mineral assemblage in reaction experiments was synthesized at 20.7 GPa and 1,600 °C for 8 h from a mixture of KR4003 natural peridotite<sup>31</sup> with an added 2.5 wt% Fe metal. In reaction runs the recovered synthetic peridotite was loaded in a second capsule, surrounded by the ATCM2 near-solidus melt composition. Additional reaction-type experiments were performed on ground mixtures of peridotite and melt compositions. In these experiments PM1 pyrolite<sup>32</sup> was used as the peridotite component and mixed with ATCM2 melt in 9:1, 7:3 and 1:1 weight ratios in Fe capsules. A single mixed experiment was performed in a Au capsule and used a starting mix of PM1:Fe:ATCM2 in 16:1:4 molar ratio.

**Experimental techniques.** High-pressure experiments were performed using a combination of end-loaded piston cylinder (3 GPa) and Walker-type multi anvil (5–21 GPa) experiments at the University of Bristol. Piston cylinder experiments employed a NaCl-pyrex assembly with a straight graphite furnace and Al<sub>2</sub>O<sub>3</sub> inner parts. Temperature was measured using type D thermocouple wires contained in an alumina sleeve and positioned immediately adjacent to the Au<sub>80</sub>Pd<sub>20</sub> sample capsule that contained the powdered starting material. We assume that the temperature gradient across the entire capsule (<2 mm) was smaller than 20 °C (refs 33, 34). The hot piston-in technique was used with a friction correction of 3% applied to the theoretical oil pressure to achieve the desired run conditions<sup>35</sup>.

Multi-anvil experiments were performed using Toshiba F-grade tungsten carbide cubes bearing 11, 8 or 4 mm truncated corners in combination with a pre-fabricated Cr-doped MgO octahedron of 18, 14 or 10 mm edge length, respectively. The relationship between oil-reservoir and sample pressure for each cell was calibrated at room and high temperature (1,200 °C) by detecting appropriate room temperature phase transitions of Bi, ZnTe and GaAs and bracketing transformations of SiO<sub>2</sub> (quartz-coesite and coesite-stishovite), Mg<sub>2</sub>SiO<sub>4</sub> (α-β and β-γ) and CaGeO<sub>3</sub> (garnet-perovskite). Calibrations are estimated to be accurate within ±1 GPa. In all experiments, desired run pressure was achieved using a slow, Eurotherm controlled, pressure ramp of ≤50 tonnes per hour. Experiments were heated after high pressure was reached with high temperatures generated using stepped graphite (18/11 cell) or straight LaCrO<sub>3</sub> furnaces (14/8 and 10/4 cells) and monitored with type C thermocouple wires. Two 10/4 experiments, performed during a period of repeated LaCrO<sub>3</sub> heater failures, used rolled 40-μm-thick Re furnaces. Temperature was quenched by turning off the furnace power before a slow decompression ramp (half the rate of experiment compression) to ambient conditions. Samples were contained in Au capsules unless temperatures exceeded its thermal stability, in which case Au<sub>80</sub>Pd<sub>20</sub> or Au<sub>75</sub>Pd<sub>25</sub> capsules were used. Run durations all exceeded 600 min and are reported in Extended Data Tables 2 and 3. Temperature uncertainties were believed to be less than ±20, 30 or 50 °C for 18/11, 14/8 and 10/4 cells respectively<sup>36,37</sup>.

Recovered samples were mounted longitudinally in epoxy, polished under oil and repeatedly re-impregnated with a low viscosity epoxy (Buehler EpoHeat) to preserve soft and water-soluble alkali carbonate components present in run products.

**Analytical techniques.** Polished and carbon-coated run products were imaged in backscatter electron mode (BSE) using a Hitachi S-3500N scanning electron microscope (SEM) with an EDAX Genesis energy dispersive spectrometer to

identify stable phases and observe product textures. Subsequently, wavelength dispersive spectroscopy (WDS) was performed using the Cameca SX100 Electron Microprobe or the Field Emission Gun Jeol JXA8530F Hyperprobe at the University of Bristol to achieve high-precision chemical analyses of run products. Analyses were performed using an accelerating voltage of 15 or 12 kV on the respective instruments, with a beam current of 10 nA. Calibrations were performed during each session using a range of natural mineral and metal standards and were verified by analysing secondary standards (as described previously<sup>6</sup>). Silicate phases were measured using a focused electron beam whereas carbonates and melts were analysed using an incident beam defocused up to a maximum size of 10 μm. Count times for Na and K were limited to 10 s on peak and 5 s on positive and negative background positions. Peak count times for other elements were 20–40 s. Additional analyses of the calcium perovskite phases grown during reaction experiments, measuring only SiO<sub>2</sub> and MgO content, were made using the Jeol instrument at 5 kV and 10 nA to ensure reported MgO contents were not influenced by secondary fluorescence from surrounding material.

The identity of experimental-produced minerals was determined using Raman spectroscopy as a fingerprint technique. Spectra were collected using a Thermo Scientific DXRxi Raman microscope equipped with an excitation laser of either 455 or 532 nm.

**Choice of bulk composition and comparison with previous studies.** Studies that investigate the alteration of oceanic crust have demonstrated that carbon incorporation does not simply occur by the addition of a single carbonate species to MORB<sup>9</sup>. It instead appears to occur by a complex amalgamation of hydrocarbon and graphite deposition related to hydrothermal fluxing above magma chambers at the mid-ocean ridge<sup>8</sup> and underwater weathering<sup>9,38–40</sup> where seawater-derived CO<sub>2</sub> reacts with leached crustal cations, often in veins. It is believed that the quantity of biotic organic carbon in the crustal assemblage is negligible compared with abiogenic organic compounds and inorganic carbonates<sup>8</sup>. These processes result in a layered crustal assemblage that, in the uppermost few hundred metres can contain up to a maximum of 4 wt% CO<sub>2</sub> in rare cases<sup>9,39</sup> but more commonly <2 wt% CO<sub>2</sub> (refs 8, 9, 39). Beneath 500 m depth the carbon content drops to between 100 and 5,000 p.p.m. CO<sub>2</sub> throughout the remainder of the 7-km-thick basaltic section<sup>8</sup>, and is mostly organic hydrocarbon species. The upper 300 m are regularly altered and can be generally thought to have compositions similar to the altered MORB rocks analysed previously<sup>41</sup>. Deeper portions of the MORB crust retain their pristine MORB compositions. It is therefore apparent that carbonated eclogite bulk compositions used in previous studies, where at least 4.4 wt% CO<sub>2</sub> was added to an eclogite by addition of ~10 wt% carbonate minerals, may not be good analogues of naturally subducting crustal sections. The compositions of these starting materials from previous studies<sup>19,42–46</sup> can be found in Extended Data Table 1. We do not include the composition of the starting material used by refs 47 or 48 as these studies were conducted in simplified chemical systems so are not directly comparable with these natural system compositions.

However, as some of the previous studies rightly identify and discuss, the composition of deeply subducted MORB is unlikely to be the same as that entering the subduction system. One process widely believed to alter the composition of downwelling MORB is sub-arc slab dehydration. Pressure (*P*)-temperature (*T*) paths of subducted slabs<sup>26</sup> can be compared with experimental studies of hydrous, carbonated and H<sub>2</sub>O-CO<sub>2</sub>-bearing eclogite compositions<sup>12,24,42,43,49</sup> and thermodynamic models<sup>11,50</sup> to conclude that slabs experience dehydration at sub-arc conditions (that is, 1–5 GPa) but will generally not reach high enough temperatures to undergo melting. Therefore, they will by and large retain their carbon components although some fraction may be lost by dissolution into aqueous fluids<sup>51,52</sup>. It is believed that sub-arc dehydration is capable of removing SiO<sub>2</sub> from the subducting assemblage, and previous carbonated MORB compositions were therefore designed to be considerably silica undersaturated (relative to fresh/ altered MORB)<sup>19,43–45</sup>. While studies<sup>53–56</sup> do indicate that SiO<sub>2</sub> can become soluble in H<sub>2</sub>O at high pressures, they infer that the solubility of silica in hydrous fluids only exceeds ~1 wt% at *T* > 900 °C at 1 GPa (higher *T* at higher *P*). In contrast, slab dehydration occurs on all prograde slab paths at *T* < 850 °C. Additionally, the composition of quenched hydrous fluids coexisting with MORB at 4 GPa and 800 °C (ref. 57) indicate that a maximum of ~12 wt% SiO<sub>2</sub> can dissolve in the fluid. Given that there should be considerably less than 10 wt% H<sub>2</sub>O (more likely << 5 wt% H<sub>2</sub>O) in subducting assemblages, this suggests a maximum SiO<sub>2</sub> loss in subducting MORB lithologies of ~0.6–1.2 wt%. The compositions used in previous studies have SiO<sub>2</sub> depletions ranging from 3 wt% up to, more commonly, 6–10 wt% SiO<sub>2</sub> relative to MORB.

We further investigated the effect of oceanic crust alteration and sub-arc dehydration on the composition of subducted MORB rocks by compiling a data set of altered MORB<sup>41</sup> and exhumed blueschist, greenschist and eclogite facies rocks from exhumed terrains worldwide to compare them with fresh MORB<sup>21</sup>, our starting



material and previous starting materials. We then assess the relevance of our starting material based on the composition of natural MORB rocks, rather than using models of the subduction process that contain few observable constraints. Results of this comparison are plotted in Extended Data Fig. 1. This analysis confirms that relative to fresh MORB, altered MORB and exhumed crustal rocks are somewhat depleted in  $\text{SiO}_2$ , up to a maximum of 6 wt%  $\text{SiO}_2$  in the most extreme case, but more commonly 0–3 wt%  $\text{SiO}_2$ . Thus, many previous starting materials are too silica undersaturated to be good analogues of subducting MORB. Furthermore, this analysis reveals that altered and exhumed MORB are not enriched in CaO compared with fresh MORB, if anything they actually contain lower CaO on average. In contrast, all previous starting materials are enriched in CaO compared with fresh MORB. This is because most previous studies introduced the carbon component to their experiment by adding ~10 wt% calcite to an eclogite-base composition. We note that SLEC1 (ref. 43) was not created in this manner, but instead this composition falls far from the MORB field as the authors used an eclogite xenolith erupted by a Hawaiian volcano as a base material. By plotting the position of the maj–cpx join, defined by the composition of our experimental phases plotted in Extended Data Fig. 5, onto Extended Data Fig. 1a, we demonstrate that our bulk composition (ATCM1), ALL-MORB<sup>21</sup>, the vast majority of the fresh MORB field, altered<sup>41</sup> and exhumed MORB samples fall on the CaO-poor side of this join, that is, on the Mg+Fe-rich side. Therefore, magnesite will be the stable carbonate phase in these compositions at high pressure (above dolomite breakdown). In contrast, all previous bulk compositions plot on the Ca-rich side of this join, or are very depleted in  $\text{SiO}_2$ , and therefore fall in a different phase field to the overwhelming majority of subducted MORB. This difference causes a considerable difference in the phase relations of our starting material relative to those used in previous studies.

We acknowledge that no single bulk composition can be a perfect analogue for the entire range of subducting MORB compositions, however, ATCM1 is a good proxy for sections of the MORB crust between ~300 m and 7 km depth that have unaltered major element compositions and low  $\text{CO}_2$  contents. Additionally, ATCM1 remains a better analogue for the uppermost portions of the MORB crust than starting materials employed in previous studies because its  $\text{CO}_2$  content is within the range of natural rocks while it is also not oversaturated in CaO or over depleted in  $\text{SiO}_2$ . This is despite it falling towards the  $\text{SiO}_2$ -rich end of the compositional spectrum of subducting MORB rocks.

**Slab  $f_{\text{O}_2}$  and carbonate survival to transition zone conditions.** Recent experiments have suggested that carbonate in eclogitic assemblages may be reduced to elemental carbon, either graphite or diamond, at depths shallower than 250 km (ref. 58). However, subducting slab geotherms are much colder than the experimental conditions investigated by this study, and additionally they are believed to contain considerable ferric iron that is further increased during de-serpentinization<sup>10</sup>. Indeed, several observations of carbonate inclusions in sub-lithospheric diamonds<sup>6,7,59</sup> require that slab carbon remains oxidized and mobile until diamond formation, far deeper than 250 km. Given the numerous observations from natural diamond samples, the general uncertainty in the mantle's  $f_{\text{O}_2}$  structure and the lack of any conclusive experimental evidence that subducting carbon becomes reduced before reaching the transition zone we posit that nearly all subducting carbon is stable as carbonate throughout the upper mantle in subducting MORB assemblages.

**Carbonated MORB melting.** Extended Data Table 2 presents the run conditions, durations and phase proportions in all carbonated MORB melting experiments, which are also summarized in Extended Data Fig. 2. Phase and melt compositions are presented in the Supplementary Tables 1–4. Phase proportions are calculated by mass balance calculations that use the mean composition of each phase as well as the reported  $1\sigma$  uncertainty in this mean as inputs. We note that the  $1\sigma$  uncertainty for some oxides in garnet and clinopyroxene minerals occasionally exceeds 1 wt%, although it is normally much smaller than this. These large uncertainties are a function of the small crystal sizes present in some runs, and not a function of sluggish reaction kinetics. Phase proportion calculations were run in a Monte Carlo loop of 10,000 calculation cycles where a varying random error was added to each oxide in each mineral phase during each iteration. Overall the distribution of varying random errors for each oxide form a Gaussian distribution with standard deviation equal to the reported  $1\sigma$  uncertainty of measurements. The reported proportions are the numerical mean of all calculation cycles and the  $r^2$  value reports the average squared sum of residuals. Low  $r^2$  values indicate that chemical equilibrium is likely to have been achieved and that mineral and melt compositions have been accurately determined.

Representative BSE images of the polished experiments are shown in Extended Data Fig. 3. Garnets in experiments at all pressures contain abundant  $\text{SiO}_2$  inclusions. In subsolidus experiments the number of inclusions increases and the definition of mineral boundaries deteriorates, which makes accurate analysis of garnet

compositions increasingly challenging. In supersolidus runs, garnet minerals adjacent, or near to, carbonatite melt pools have well defined edges and contain fewer inclusions. However, far from quenched melts the textures of garnets remain small and pervasively filled with inclusions, indicating the influence of melt fluxing on mineral growth. With increasing pressure, garnets become increasingly majoritic, with increasing quantities of octahedral silicon.

Clinopyroxene was observed in all subsolidus experiments, as euhedral crystals that are often spatially associated with the carbon-bearing phase. Cpx abundance falls with increasing pressure and their compositions becoming increasingly dominated by sodic components (jadeite, aegirine and  $\text{NaMg}_{0.5}\text{Si}_{2.5}\text{O}_6$ ) at high pressure (Extended Data Fig. 5). Cpx only disappears from the stable phase assemblage in supersolidus experiments at 20.7 GPa.  $\text{SiO}_2$  is observed in all runs and are small, often elongated tabular-shaped crystals. An oxide, either  $\text{TiO}_2$  at low pressure or an Fe–Ti oxide above 13 GPa (as described previously<sup>24</sup>) are observed in all subsolidus runs.

The carbon-bearing phase in subsolidus experiments changes with increasing pressure. At 3 GPa  $\text{CO}_2$ , marked by the presence of voids in the polished sample, is stable. This converts to dolomite at 7.9 GPa, consistent with the position of the reaction  $2\text{cs} + \text{dol} = \text{cpx} + \text{CO}_2$  (ref. 22). Beyond ~9 GPa dolomite becomes unstable and breaks down into magnesite + aragonite<sup>23</sup>. Therefore, because the ATCM1 bulk composition lies on the Mg+Fe<sup>2+</sup>-rich side of the garnet–cpx join (Extended Data Figs 1a and 5), magnesite replaces dolomite as the carbon host in the experimental phase assemblage. This differs from experiments in previous studies, where aragonite was dominant because bulk compositions fall on the opposite side of the garnet–cpx join. It is clear from the ternary diagrams (Extended Data Fig. 5) that while the tie-line between garnet and cpx remains, magnesite and aragonite cannot coexist in a MORB bulk composition. Finally, at pressures above 15 GPa, Na-carbonate becomes stable in the subsolidus phase assemblage. This is chemographically explained by the rotation of the garnet–cpx tie-line with increasing pressure (EDF5). Its appearance can also be justified as a necessary host of sodium at increasing pressure, since aside from clinopyroxene there is no other Na-rich phase stable on the Mg+Fe side of the maj–cpx join.

The appearance of silicate melt, containing dissolved  $\text{CO}_2$  (estimated by difference), defines the solidus at 3 GPa. This may initially appear to contradict the results of some previous studies, which find carbonatite melts are produced near the solidus of carbonated eclogite at pressures lower than 7 GPa (refs 43, 45, 46). However, this is easily explained by the differences in  $\text{CO}_2$  and  $\text{SiO}_2$  content used in these studies. The higher  $\text{CO}_2$  and lower  $\text{SiO}_2$  contents of previous studies stabilize carbonate melt to lower temperatures relative to silicate melts. Indeed, we note that our results are consistent with those described previously<sup>42,44</sup> (the two previous studies with the least depleted  $\text{SiO}_2$ ), which also observed that near-solidus melts below 5 GPa were basaltic to dacitic silicate melts containing dissolved  $\text{CO}_2$ . The results of one paper<sup>19</sup> are not entirely self-consistent, in that at some pressures between 3.5 and 5.5 GPa the authors observed silicate melts before carbonate melts (4.5 and 5 GPa), whereas this relationship is sometimes reversed (5 GPa in AuPd capsules) or both melts were observed together (3.5 GPa). The observation of two immiscible melts in previous studies probably reflects the maximum  $\text{CO}_2$  solubility in silicate melts. Since our bulk composition has less  $\text{CO}_2$ , akin to natural rocks, we do not observe liquid immiscibility.

In all experiments above 7 GPa, near-solidus melt compositions are carbonatitic and essentially silica-free. This result is notably different from those described previously<sup>19</sup>, which reported that near-solidus melts were a mixture of silicate, carbonated silicate and carbonatite melts. We believe this contrast is caused by the interpretation of experimental run textures. Whereas ref. 19 identified regions of fine-grained material consisting of mixtures of stable phases from elsewhere in the capsule as quenched melts, we have not followed the same interpretation of these features. Although we do recognize similar features in some run products, we have interpreted these features as a consequence of poor crystal growth in regions far from the influence of melt fluxing. In all supersolidus experiments, we observed regions of carbonatite material (typically <1 wt%  $\text{SiO}_2$ ) that is fully segregated from surrounding silicate minerals and possesses a typical carbonate-melt quench texture (Extended Data Fig. 3). Silicate minerals in close proximity to these melt pools are larger than those elsewhere in the same experiment, have well-defined crystal boundaries and contain few inclusions. Therefore, we attribute the variable texture and regions of fine-grained material present in experiments to the location of melt within experiments, which has a tendency to segregate to isolated regions of capsules under influence of temperature gradients. Although melt segregation occurs in all supersolidus experiments, the efficiency of segregation and size of melt pools considerably increases with rising temperature above the solidus. Extended Data Figure 4 shows the highly systematic evolution of the melt compositions reported from our study with increasing pressure, strongly supporting our interpretations.

Carbonatite melts are calcic, Ca number > 0.5 (Ca number =  $\text{Ca}/[\text{Ca}+\text{Mg}+\text{Fe}]$ ), despite subsolidus carbonates being dominated by magnesite (Extended Data Fig. 4). Melts have high concentrations of  $\text{TiO}_2$  (typically 1–3.5 wt%),  $\text{P}_2\text{O}_5$  (0.4–1.5 wt%) and  $\text{K}_2\text{O}$  (0.3–1.5 wt%) and a variable Mg number (0.33–0.7 defined as  $\text{Mg}/[\text{Mg}+\text{Fe}]$ ). The alkali content of melts, strongly dominated by  $\text{Na}_2\text{O}$  due to the bulk composition, increases with pressure (from 1 to ~15 wt%  $\text{Na}_2\text{O}$  at 7.9 and 20.7 GPa respectively; Extended Data Fig. 4). This increasing  $\text{Na}_2\text{O}$  content is driven by the decreasing compatibility of  $\text{Na}_2\text{O}$  in the residual mantle phase assemblages as the abundance of stable clinopyroxene falls. At 20.7 GPa the melt composition, as evidenced both by constant phase proportions and consistent melt/majorite compositions, remains constant over a temperature interval of ~350 °C above the solidus. It is only when temperature reaches 1,530–1,600 °C (runs #16 and #31) that the silica content of the melt begins to increase (to 8.7 wt%) and  $\text{CO}_2$  content falls as melts start to become silica-carbonatites.

One experiment (#33) aimed to verify that measured low-degree melt compositions are accurate, and are not affected by analytical problems related to the small size of melt pools, was conducted at 20.7 GPa. In this experiment the abundance of carbonate melt was increased by adding a mix replicating the low degree melt composition ATCM2 to ATCM1 in a mass ratio of 1:3. If the composition of low-degree melts has been accurately determined in 'normal' experiments then this addition will have a negligible effect on phase relations or the compositions of the garnet,  $\text{SiO}_2$  or melt; it would simply increase the melt abundance. The result of this experiment has a similar texture to all other experiments, where carbonatite melt segregates to one end of the capsule and is adjacent to large, well-formed majoritic garnets. The far end of the capsule has a much smaller crystal size, crystals have ragged edges, garnets are full of inclusions and  $\text{SiO}_2$  is present along grain-boundaries and triple junctions (Extended Data Fig. 3h). Mineral and melt compositions, although not exactly identical, are similar to those measured in 'normal' experiments (to achieve identical compositions an iterative approach would be required that was not deemed to be necessary) thus confirming that near-solidus melt compositions have been accurately determined. The presence of fine-grained material away from segregated melt also acts to further confirm our hypothesis regarding the vital importance of melt presence for growing large crystals during experiments.

**Subsolidus carbonate species at high pressure.** Comparing our starting material and results with those of previous studies using ternary and quaternary projections (Extended Data Fig. 5) reveals that it is not possible for both magnesite and aragonite to coexist alongside majorite and clinopyroxene owing to stable mineral phase fields (see earlier). Thus, in Mg-Fe-dominated compositions, such as our starting material, magnesite is the stable carbonate at high-pressure subsolidus conditions. Whereas in Ca-dominated compositions aragonite will be the stable carbonate beyond the pressure of dolomite dissociation. Natural subducting MORB compositions, which contain, at most, a similar quantity of  $\text{CO}_2$  to our bulk composition<sup>11</sup>, almost all lie on the Ca-poor side of the majorite-clinopyroxene join (Extended Data Figs 1 and 5). In this situation, as our experiments demonstrate, cpx remains an important Na-host in MORB assemblages to high pressures alongside  $[\text{Na},\text{K}]_{0.33}\text{Ca}_{0.67}\text{CO}_3$  structured carbonate. Ca-rich compositions containing subsolidus  $\text{CaCO}_3$  experience different phase relations because aragonite can dissolve considerable  $\text{Na}_2\text{O}$  and so is the sole Na-host in these compositions. We conclude that because the majority of natural MORB rocks fall on the Mg+Fe side of the maj-cpx join, like our bulk composition, that the phase relations determined in this study are applicable to the case of natural subduction. Therefore, the melting point depression we observe along the carbonated MORB solidus at uppermost transition zone pressures is generally applicable to subducted oceanic crust.

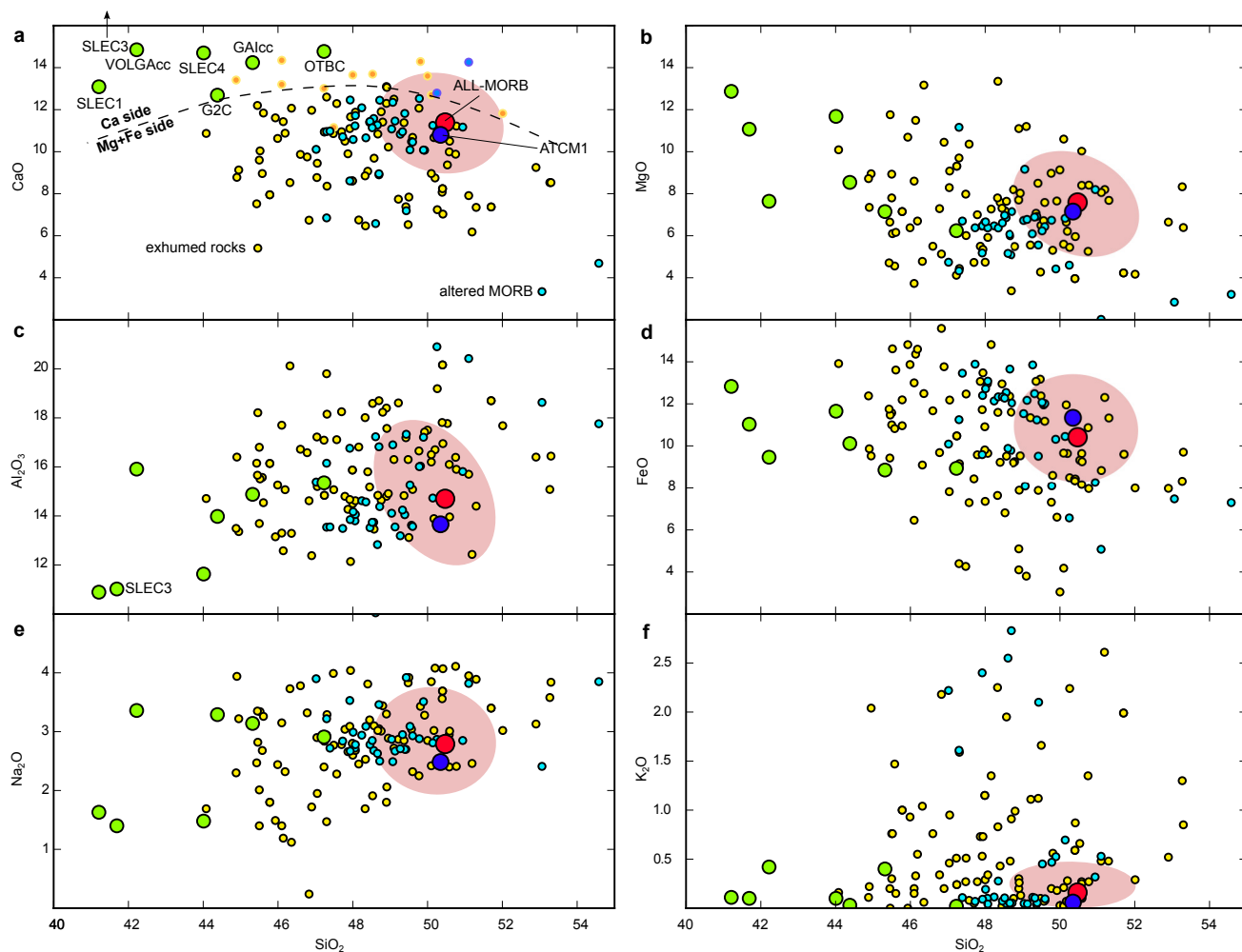
**Melt-mantle reactions.** Without the influence of slab-derived melts, the anhydrous transition zone peridotite assemblage at 20.7 GPa and 1,600 °C (experiment G168 and G176) is dominated by Na-poor majorite and wadsleyite (Mg number = 0.90) (Extended Data Fig. 6, Extended Data Table 3 and Supplementary Table 5a). Upon reaction with the near-solidus alkaline carbonatite defined during melting experiments, ATCM2, a clearly defined reaction zone is observed between this ambient peridotite assemblage and the infiltrating melt (Extended Data Fig. 6). The products of this reaction are garnet containing a notable  $\text{Na}_2\text{X}^{2+}\text{Si}_5\text{O}_{12}$  majorite component,  $\text{Ca}(\text{Si},\text{Ti})\text{O}_3$  perovskite, ringwoodite, ferropericlase and diamond. All of these phases were identified using Raman spectroscopy (Extended Data Fig. 7) and their compositions are presented in Supplementary Table 5a. Raman spectroscopy alone, which was performed before any sample polishing using diamond-based products, confirms the creation of diamond during these reactions. We have not observed diamond using SEM techniques and believe that it resides as sub-micrometre-sized inclusions in the various reaction-product minerals where it is seen by spectroscopic methods. The experiments performed on

intimately mixed powders of melt and pyrolite also form the same phase assemblages (Extended Data Table 3) and mineral compositions from those runs are also presented in Supplementary Table 5b, c.

We observed the reaction products as new crystals floating in the residual carbonatite melt and/or nucleated on the relics of the peridotite assemblage, thus creating zoned minerals. We have demonstrated that the composition of majorite minerals crystallizing during the reactions lie between those expected for peridotitic and eclogitic minerals at a similar pressure and possibly explain intermediate-composition diamond-hosted majorites (Fig. 2). We suggest that the full range of intermediate inclusion compositions might be created by the gradual shift in phase compositions, from those we observe towards more peridotitic minerals as the melt composition reacts with increasing quantities of mantle material. Additionally we have shown that the compositions of calcium perovskite (Extended Data Fig. 8) and ferropericlase (Fig. 3) formed during the reactions are consistent with diamond-hosted minerals of those species. Further experiments, across the solidus ledge and into the uppermost lower mantle pressure range are required to test whether melt-mantle interactions account for all diamond-hosted inclusions.

- Walter, M. J. Melting of garnet peridotite and the origin of komatiite and depleted lithosphere. *J. Petrol.* **39**, 29–60 (1998).
- Walter, M. J., Nakamura, E., Trönnnes, R. G. & Frost, D. J. Experimental constraints on crystallization differentiation in a deep magma ocean. *Geochim. Cosmochim. Acta* **68**, 4267–4284 (2004).
- Watson, E., Wark, D., Price, J. & Van Orman, J. Mapping the thermal structure of solid-media pressure assemblies. *Contrib. Mineral. Petrol.* **142**, 640–652 (2002).
- Schilling, F. & Wunder, B. Temperature distribution in piston-cylinder assemblies: numerical simulations and laboratory experiments. *Eur. J. Mineral.* **16**, 7–14 (2004).
- McDade, P. et al. Pressure corrections for a selection of piston-cylinder cell assemblies. *Mineral. Mag.* **66**, 1021–1028 (2002).
- Walter, M. J., Thibault, Y., Wei, K. & Luth, R. W. Characterizing experimental pressure and temperature conditions in multi-anvil apparatus. *Can. J. Phys.* **73**, 273–286 (1995).
- Hernlund, J., Leinenweber, K., Locke, D. & Tyburczy, J. A. A numerical model for steady-state temperature distributions in solid-medium high-pressure cell assemblies. *Am. Mineral.* **91**, 295–305 (2006).
- Nakamura, K. & Kato, Y. Carbonatization of oceanic crust by the seafloor hydrothermal activity and its significance as a  $\text{CO}_2$  sink in the Early Archean. *Geochim. Cosmochim. Acta* **68**, 4595–4618 (2004).
- Coogan, L. a. & Gillis, K. M. Evidence that low-temperature oceanic hydrothermal systems play an important role in the silicate-carbonate weathering cycle and long-term climate regulation. *Geochim. Geophys. Geosyst.* **14**, 1771–1786 (2013).
- Coogan, L. a. & Dosso, S. E. Alteration of ocean crust provides a strong temperature dependent feedback on the geological carbon cycle and is a primary driver of the Sr-isotopic composition of seawater. *Earth Planet. Sci. Lett.* **415**, 38–46 (2015).
- Kelley, K. A., Plank, T., Ludden, J. & Staudigel, H. Composition of altered oceanic crust at ODP Sites 801 and 1149. *Geochim. Geophys. Geosyst.* **4**, 6 (2003).
- Hammouda, T. High-pressure melting of carbonated eclogite and experimental constraints on carbon recycling and storage in the mantle. *Earth Planet. Sci. Lett.* **214**, 357–368 (2003).
- Dasgupta, R., Hirschmann, M. M. & Withers, A. C. Deep global cycling of carbon constrained by the solidus of anhydrous carbonated eclogite under upper mantle conditions. *Earth Planet. Sci. Lett.* **227**, 73–85 (2004).
- Yaxley, G. M. & Green, D. H. Experimental demonstration of refractory carbonate-bearing eclogite and siliceous melt in the subduction regime. *Earth Planet. Sci. Lett.* **128**, 313–325 (1994).
- Dasgupta, R., Hirschmann, M. M. & Dellas, N. The effect of bulk composition on the solidus of carbonated eclogite from partial melting experiments at 3 GPa. *Contrib. Mineral. Petrol.* **149**, 288–305 (2005).
- Gerbode, C. & Dasgupta, R. Carbonate-fluxed melting of MORB-like pyroxenite at 2.9 GPa and genesis of HIMU ocean island basalts. *J. Petrol.* **51**, 2067–2088 (2010).
- Litasov, K. & Ohtani, E. The solidus of carbonated eclogite in the system  $\text{CaO-Al}_2\text{O}_3\text{-MgO-SiO}_2\text{-Na}_2\text{O-CO}_2$  to 32 GPa and carbonatite liquid in the deep mantle. *Earth Planet. Sci. Lett.* **295**, 115–126 (2010).
- Keshav, S. & Gudfinnsson, G. H. Experimentally dictated stability of carbonated oceanic crust to moderately great depths in the Earth: results from the solidus determination in the system  $\text{CaO-MgO-Al}_2\text{O}_3\text{-SiO}_2\text{-CO}_2$ . *J. Geophys. Res.* **115**, B05205 (2010).
- Molina, J. F. & Poli, S. Carbonate stability and fluid composition in subducted oceanic crust: an experimental study on  $\text{H}_2\text{O-CO}_2$  bearing basalts. *Earth Planet. Sci. Lett.* **176**, 295–310 (2000).
- Connolly, J. A. D. Computation of phase equilibria by linear programming: a tool for geodynamic modeling and its application to subduction zone decarbonation. *Earth Planet. Sci. Lett.* **236**, 524–541 (2005).
- Manning, C. E. Geochemistry: a piece of the deep carbon puzzle. *Nature Geosci.* **7**, 333–334 (2014).

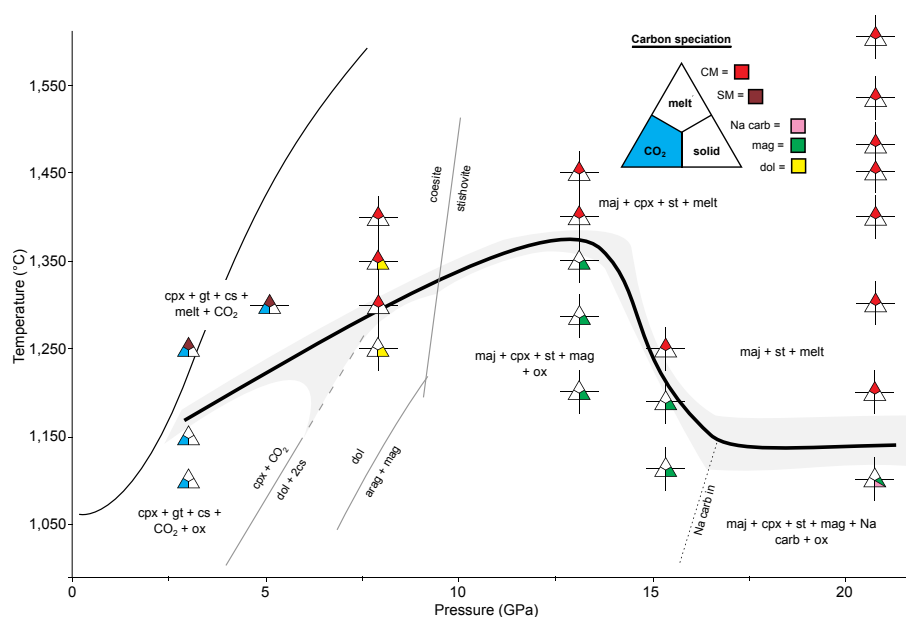
52. Ague, J. J. & Nicolescu, S. Carbon dioxide released from subduction zones by fluid-mediated reactions. *Nature Geosci.* **7**, 355–360 (2014).
53. Manning, C. E. The chemistry of subduction-zone fluids. *Earth Planet. Sci. Lett.* **223**, 1–16 (2004).
54. Newton, R. C. & Manning, C. E. Quartz solubility in H<sub>2</sub>O–NaCl and H<sub>2</sub>O–CO<sub>2</sub> solutions at deep crust-upper mantle pressures and temperatures: 2–15 kbar and 500–900 °C. *Geochim. Cosmochim. Acta* **64**, 2993–3005 (2000).
55. Newton, R. C. & Manning, C. E. Solubility of enstatite + forsterite in H<sub>2</sub>O at deep crust/upper mantle conditions: 4 to 15 kbar and 700 to 900 °C. *Geochim. Cosmochim. Acta* **66**, 4165–4176 (2002).
56. Newton, R. C. & Manning, C. E. Thermodynamics of SiO<sub>2</sub>–H<sub>2</sub>O fluid near the upper critical end point from quartz solubility measurements at 10 kbar. *Earth Planet. Sci. Lett.* **274**, 241–249 (2008).
57. Kessel, R., Schmidt, M., Ulmer, P. & Pettke, T. Trace element signature of subduction-zone fluids, melts and supercritical liquids at 120–180 km depth. *Nature* **437**, 724–727 (2005).
58. Stagno, V. *et al.* The oxygen fugacity at which graphite or diamond forms from carbonate-bearing melts in eclogitic rocks. *Contrib. Mineral. Petrol.* **169**, 16 (2015).
59. Zedgenizov, D. A., Kagi, H., Shatsky, V. S. & Ragozin, A. L. Local variations of carbon isotope composition in diamonds from São-Luis (Brazil): evidence for heterogenous carbon reservoir in sublithospheric mantle. *Chem. Geol.* **363**, 114–124 (2014).



**Extended Data Figure 1 | Comparison of experimental compositions with natural rocks. a–f,** ‘Fresh’ MORB rocks (red field), ALL-MORB<sup>21</sup> (red circle), altered MORB rocks<sup>41</sup> (pale blue circles), exhumed blueschist, greenschist and/or eclogitic rocks (yellow circles) and starting material from this (dark blue circle) and previous studies (green circles) of carbonated MORB compositions. In **a**, rocks altered MORB and exhumed rock compositions that fall on the Mg-Fe side of the maj-cpx join from

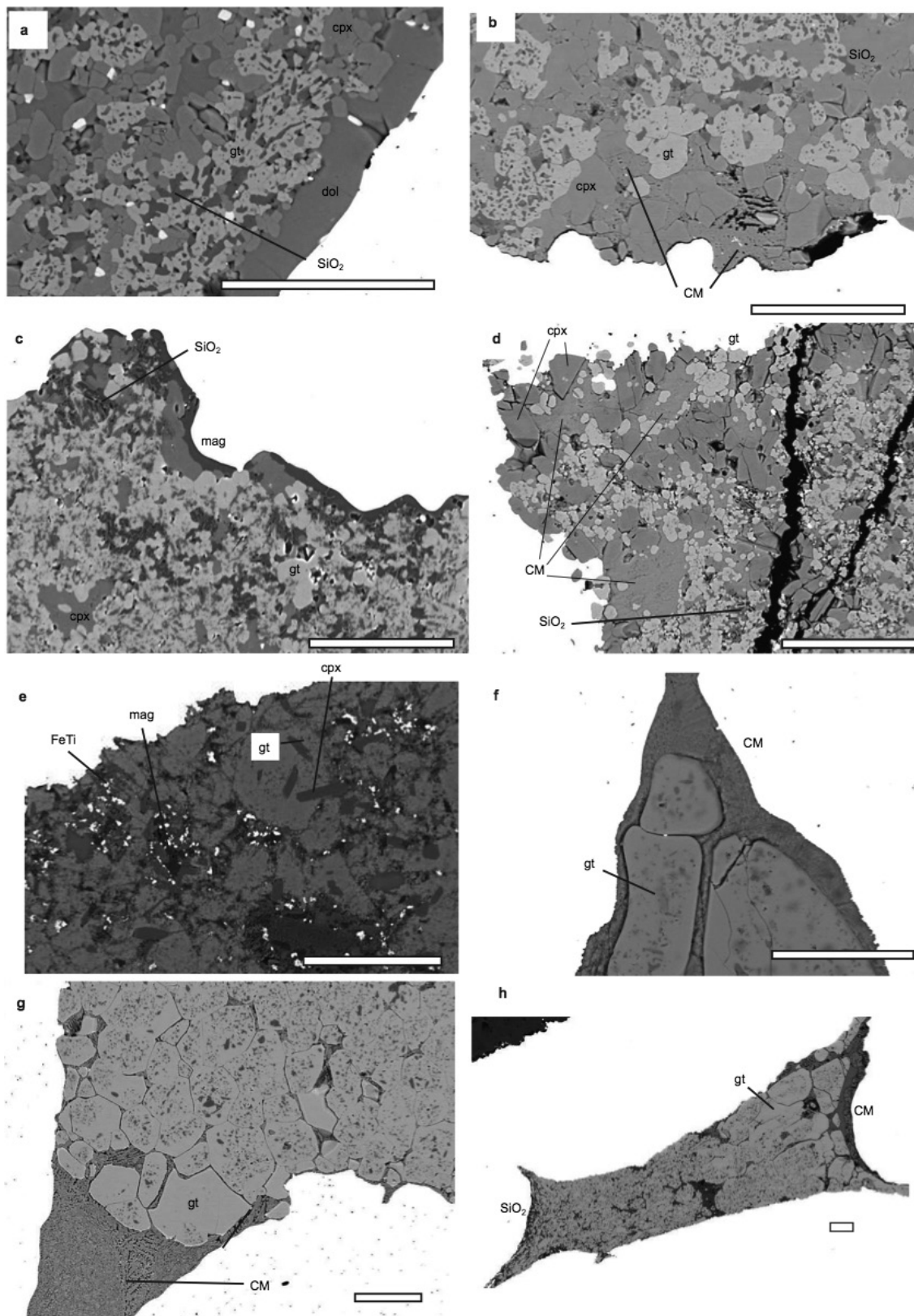
Extended Data Fig. 5 plot below the dashed line, compositions that lie on the Ca side of this join are plotted as orange circles with yellow outlines or purple circles with blue outlines and sit above the dashed curve. This confirms that magnesite will be the stable carbonate phase at high pressure in the vast majority of natural crustal rocks, as is the case for ATCM1. Data and corresponding references for this figure are provided in the online source data file.





**Extended Data Figure 2 | Experimental results/phase diagram and interpreted solidus position.** The reactions clinopyroxene +  $\text{CO}_2$  = dolomite + 2coesite and dolomite = magnesite + aragonite are from refs 22 and 23 respectively. The upper left curve is the anhydrous MORB solidus. Note that due to temperature gradients in experiments

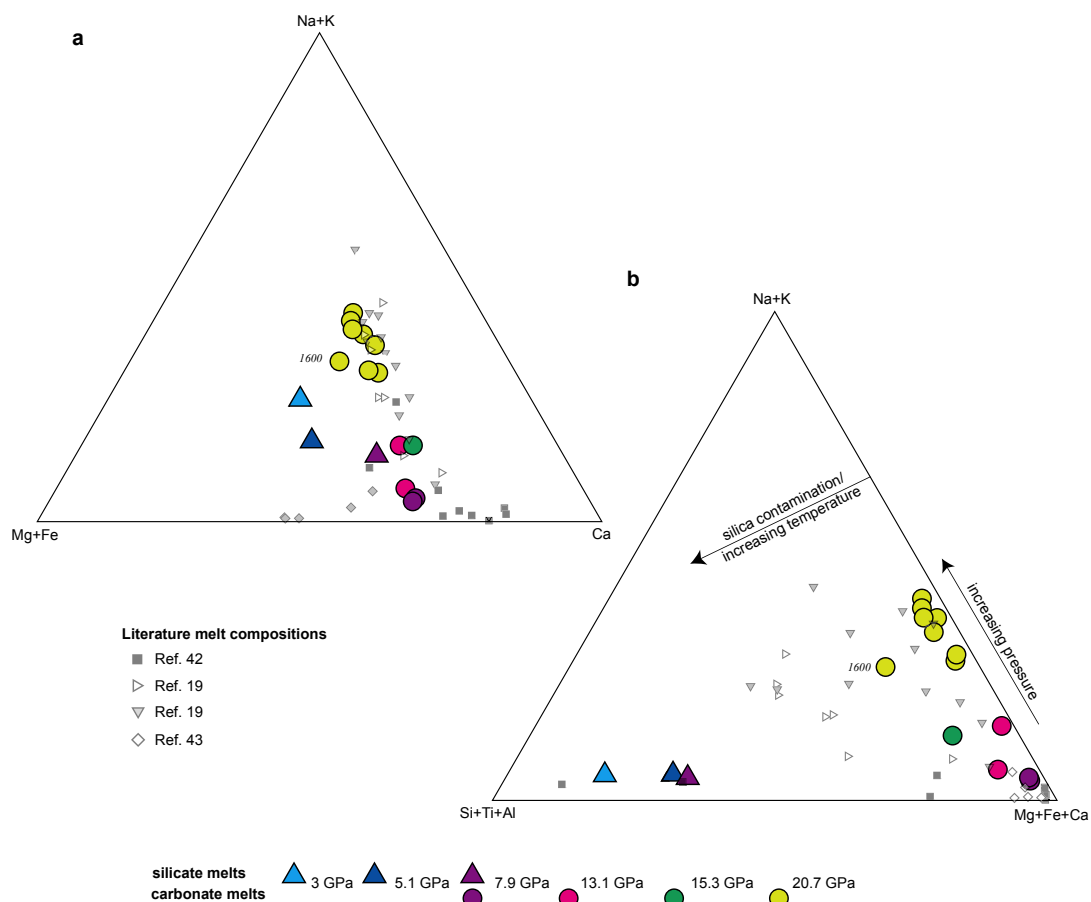
at 8 GPa, a small quantity of dolomite is observed coexisting with melt in one experiment above the solidus, present at the cold end of the capsule. arag, aragonite; CM, carbonatite melt; cpx, clinopyroxene; cs, coesite; dol, dolomite; gt, garnet; mag, magnesite; maj, majoritic garnet; Na carb, Na carbonate; ox, FeTi oxide; SM, silicate melt; st, stishovite.



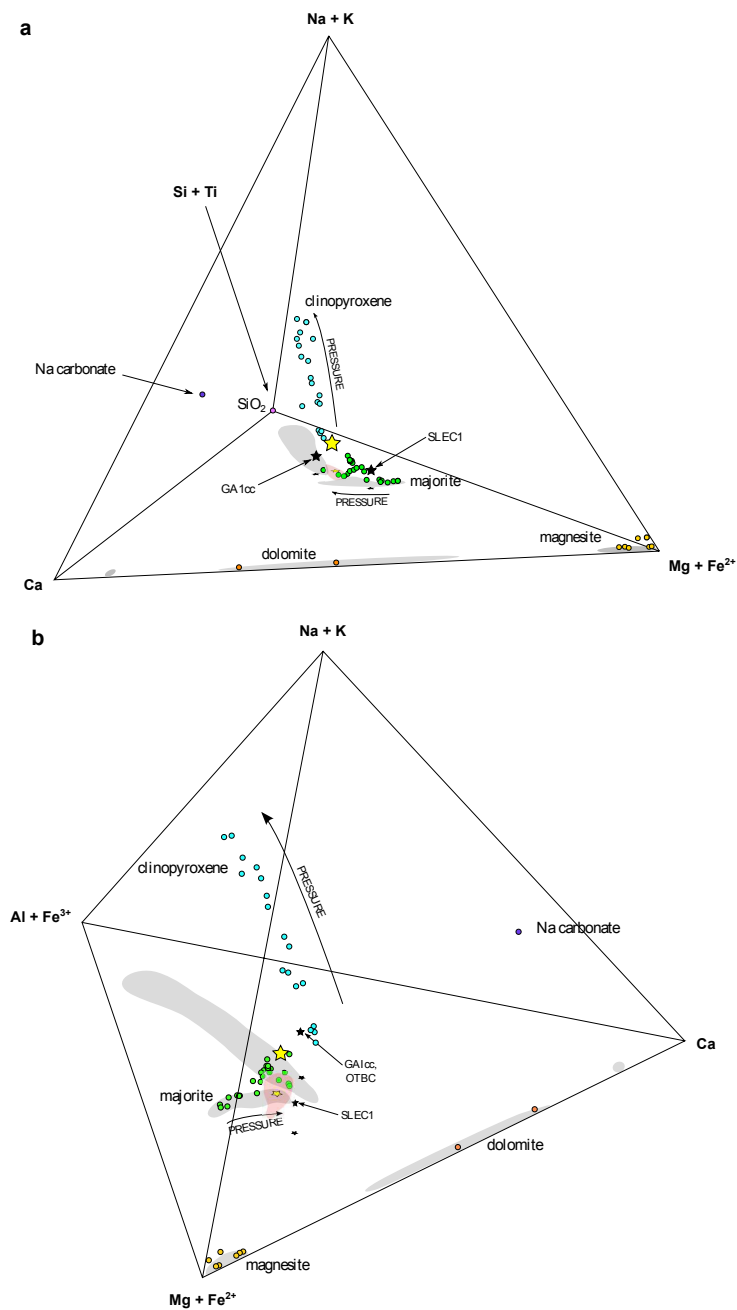
**Extended Data Figure 3 | BSE images of experimental products.**

**a**, 7.9 GPa, 1,250 °C; **b**, 7.9 GPa, 1,350 °C; **c**, 13.1 GPa, 1,350 °C; **d**, 13.1 GPa, 1,450 °C; **e**, 20.7 GPa, 1,100 °C; **f**, 20.7 GPa, 1,480 °C; **g**, 20.7 GPa, 1,600 °C;

**h**, sandwich experiment, 20.7 GPa, 1,400 °C. Scale bars, 10 μm. CM, carbonatite melt; cpx, clinopyroxene; dol, dolomite; FeTi, FeTi oxide; gt, garnet; mag, magnesite.



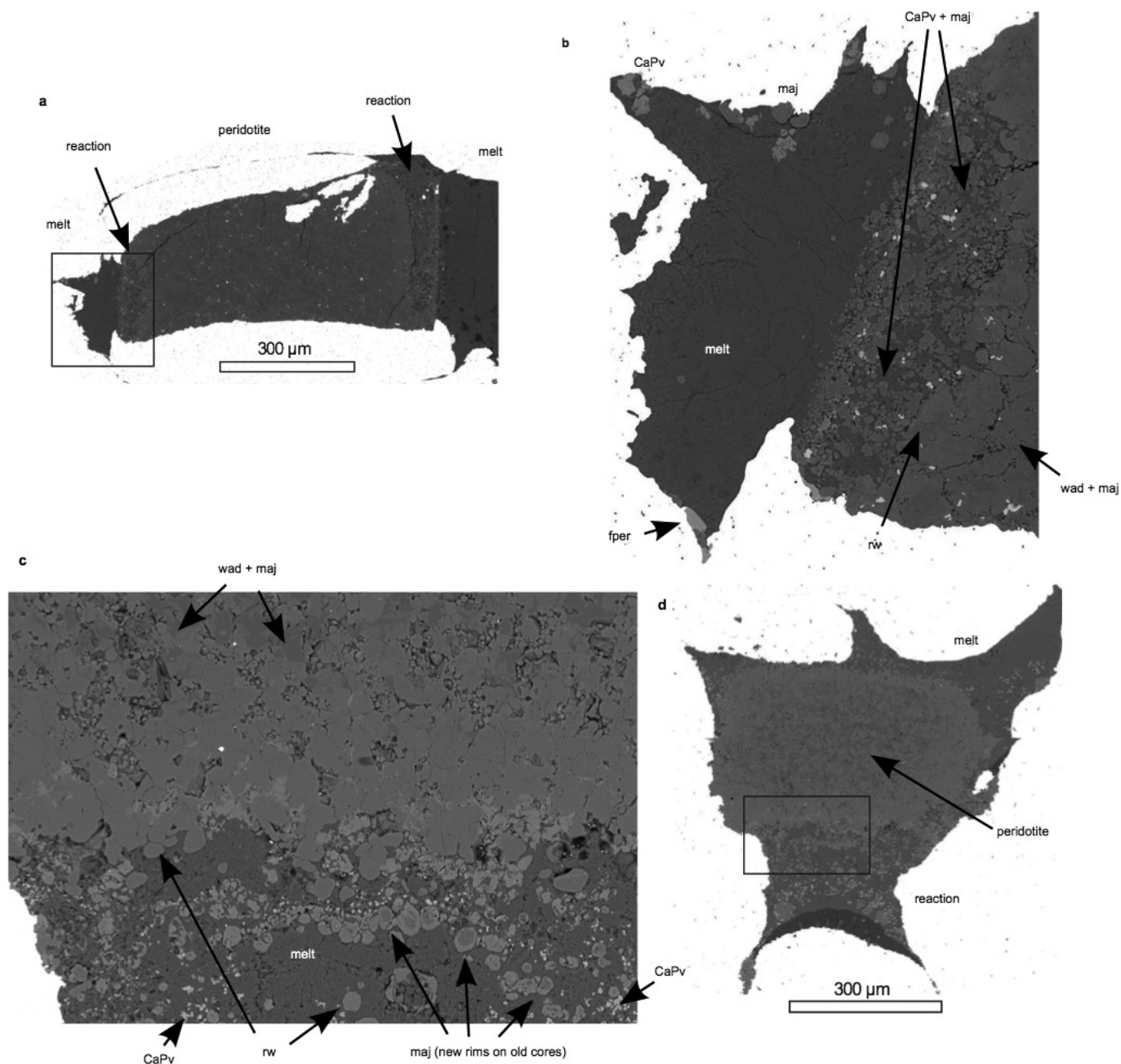
**Extended Data Figure 4 | Composition of experimental melts from this study. a, b,** Experimental melts from selected previous studies marked with semi-transparent greyscale symbols. **b,** The effects of increasing pressure, temperature and the effect of contamination due to partial analysis of silicate minerals surrounding small melt pools are shown.



**Extended Data Figure 5 | The composition of experimental phases from this study projected into two quaternary plots. a, b, [Ca]-[Mg+Fe<sup>2+</sup>]-[Si+Ti]-[Na+K] (a) and [Mg+Fe<sup>2+</sup>]-[Ca]-[Al+Fe<sup>3+</sup>]-[Na+K] (b). In both diagrams the grey fields are the compositional data projected onto the basal ternary. The red field is the range of natural MORB compositions**

projected onto the basal ternary. The yellow star plotted in the four-component system and projected onto the basal ternary is ATCM1 (our bulk composition) while the black stars are bulk compositions from previous studies<sup>25-27</sup>.

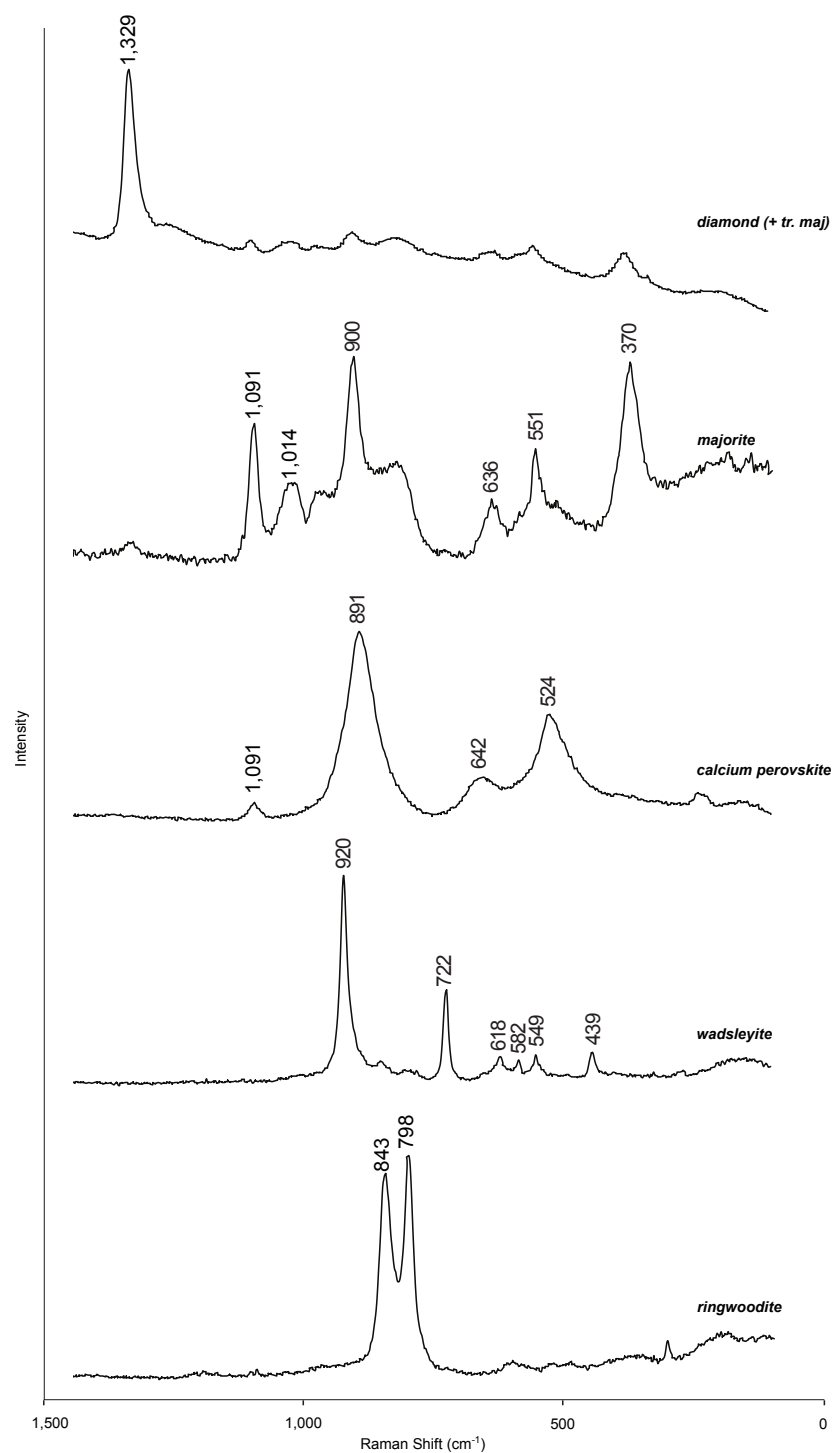




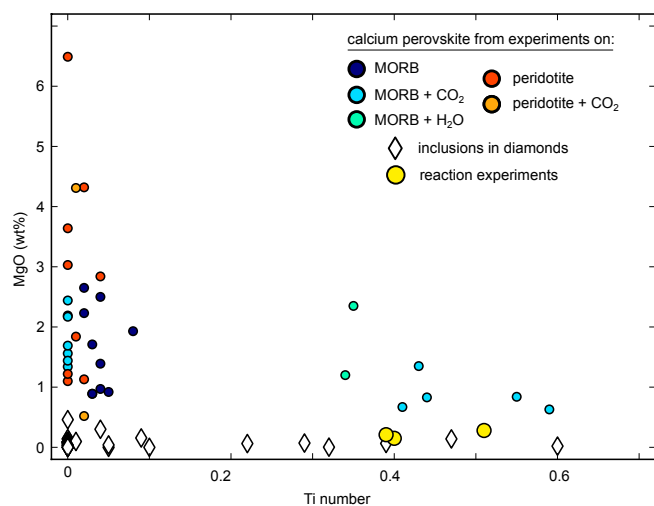
#### Extended Data Figure 6 | BSE images of reaction experiments.

**a–d**, G169 (**a**, **b**) and G177 (**c**, **d**). In both experiments a reaction zone and remaining carbonatite melt surrounds the unreacted peridotite region. **a**, An overview of G169. **b**, A close up of the reaction in G169 containing newly crystallized calcium perovskite, majorite, ferropericlase

and ringwoodite minerals. **c**, A close up of the reaction products in G177, which consist of small bright calcium perovskites, new majorite that is often observed as a rim on relic peridotitic garnet and ringwoodite. **d**, An overview of G177. CaPv, calcium perovskite; fper, ferropericlase; maj, majorite; rw, ringwoodite; wad, wadsleyite.



**Extended Data Figure 7 | Raman spectra of minerals from reaction experiment G177 measured using a blue  $455\text{ cm}^{-1}$  excitation laser.** The position of the main peaks in each collected spectrum have been labelled with their shift from the excitation laser in  $\text{cm}^{-1}$ .



**Extended Data Figure 8 | Comparison of diamond-hosted calcium perovskite inclusions with experimental mineral compositions in MgO versus Ti number space.** Ti number =  $\text{Ti}/[\text{Ca} + \text{Ti}]$ . Data and corresponding references for this figure are provided in the online source data file.

Extended Data Table 1 | Starting materials used in this and previous studies

	ALL MORB <sup>21</sup>	IODP 1256 MORB <sup>20</sup>	ATCM1 (this study)	SLEC1 <sup>43</sup>	OTBC <sup>42</sup>	GA1cc <sup>19,41</sup>	VOLGAcc <sup>19</sup>	SLEC2 <sup>45</sup>	SLEC3 <sup>45</sup>	SLEC4 <sup>45</sup>	G2C <sup>46</sup>	ATCM2 (this study)	KR4003 <sup>31</sup>	PM1 <sup>32</sup>
SiO <sub>2</sub>	50.47	51.48	50.35	41.21	47.23	45.32	42.22	30.29	41.69	44.01	44.38	0.71	44.90	45.74
TiO <sub>2</sub>	1.68	1.44	1.33	2.16	-	1.34	1.43	1.58	2.18	2.31	1.75	1.52	0.16	0.21
Al <sub>2</sub> O <sub>3</sub>	14.70	14.18	13.66	10.89	15.35	14.88	15.91	7.95	11.02	11.63	13.98	0.33	4.26	5.01
Cr <sub>2</sub> O <sub>3</sub>													0.41	0.27
FeO	10.43	11.90	11.35	12.83	8.93	8.85	9.46	13.87	11.03	11.65	10.11	8.43	8.02	8.07
MnO	0.18	0.22	0.21	0.12	-	0.15	0.14	0.23	0.12	0.13	0.25	-	0.13	0.14
MgO	7.58	7.30	7.15	12.87	6.24	7.15	7.64	14.28	11.07	11.68	8.54	6.50	37.30	34.57
CaO	11.39	10.78	10.80	13.09	14.77	14.24	14.85	14.88	16.89	14.70	12.69	23.21	3.45	3.86
Na <sub>2</sub> O	2.79	2.53	2.48	1.63	2.91	3.14	3.36	1.75	1.40	1.48	3.29	16.35	0.22	0.59
K <sub>2</sub> O	0.16	0.06	0.06	0.11	0.02	0.40	0.42	0.13	0.10	0.10	0.03	0.57	0.09	0.08
P <sub>2</sub> O <sub>5</sub>	0.18	0.11	0.10			0.14	0.15					0.37		
NiO													0.24	0.20
CO <sub>2</sub>			2.52	5.00	4.43	4.40	4.40	14.99	4.42	2.21	5.00	42.01		
Total	99.57	100.00	100.00	99.91	99.88	100.01	99.98	99.95	99.92	99.90	100.02	100.00	98.18	98.74
Ca#	0.38	0.36	0.36	0.32	0.48	0.46	0.45	0.33	0.41	0.37	0.39	0.60	0.06	0.07
Mg#	0.57	0.52	0.53	0.64	0.56	0.59	0.59	0.65	0.64	0.64	0.60	0.58	0.89	0.88

Ca number = Ca/[Ca+Mg+Fe]. Mg number = Mg/[Mg+Fe].



Extended Data Table 2 | Summary of run conditions and products for carbonated MORB melting experiments

	<i>P</i> (GPa)	<i>T</i> (°C)	<i>time</i> (min)	<i>capsule</i>	<i>run products</i>	<i>comments</i>	<i>gt</i>	<i>cpx</i>	<i>cs/st</i>	<i>ox</i>	<i>dol</i>	<i>mag</i>	<i>Na carb</i>	<i>SM</i>	<i>CM</i>	<i>CO<sub>2</sub></i>	<i>r</i> <sup>2</sup>
<b>162G</b>	3	1100	7200	Au <sub>80</sub> Pd <sub>20</sub>	gt, cpx, cs, rut, CO <sub>2</sub>		29.96	59.02	7.84	0.24						2.52	<b>0.12</b>
<b>161G</b>	3	1150	7200	Au <sub>80</sub> Pd <sub>20</sub>	gt, cpx, cs, rut, CO <sub>2</sub>		27.69	61.07	7.87	0.20						2.52	<b>0.16</b>
<b>#5</b>	3	1250	2880	Au <sub>80</sub> Pd <sub>20</sub>	gt, cpx, cs, rut, SM, CO <sub>2</sub>		28.01	53.93	5.76	0.08				9.71		2.03	<b>0.55</b>
<b>#32</b>	5.1	1300	1560	Au <sub>80</sub> Pd <sub>20</sub>	gt, cpx, cs, rut, SM		30.22	40.81	7.92					20.97			<b>0.20</b>
<b>#9</b>	7.9	1250	1470	Au	maj, cpx, cs, rut, dol	dolomite segregated at one end	50.39	32.05	12.07	0.83	5.48						<b>0.18</b>
<b>#10</b>	7.9	1300	1440	Au	maj, cpx, cs, rut, CM	melt in centre (~25 °C above TC)	45.43	34.87	12.53	0.51					6.57		<b>0.12</b>
<b>#14</b>	7.9	1350	1770	Au <sub>75</sub> Pd <sub>25</sub>	maj, cpx, cs, rut, dol, CM	melt adjacent to TC, dolomite at cold end (~50 °C below TC)	44.49	34.96	13.62	0.63	1.95				4.42		<b>0.11</b>
<b>#34</b>	7.9	1400	1440	PtRe	maj, cpx, cs, CM	Fe loss to capsule											
<b>#29</b>	13.1	1200	1440	Au	maj, cpx, st, rut, mag	magnesite segregated at one end	64.37	14.85	15.31	0.49		4.75					<b>0.39</b>
<b>#40</b>	13.1	1280	120	Au	maj, cpx, st, rut, mag	Experiment failed after 2 hours	63.45	16.14	14.86	0.40		4.82					<b>1.02</b>
<b>#12</b>	13.1	1350	1440	Au	maj, cpx, st, rut, mag	magnesite segregated at one end	63.66	16.20	15.66	0.11		4.13					<b>3.70</b>
<b>#39</b>	13.1	1350	1560	Au	maj, cpx, st, rut, mag	repeat of #12	62.88	15.36	15.88	0.52		4.33					<b>2.89</b>
<b>#17</b>	13.1	1400	1440	Au	maj, cpx, st, CM	very small melt pools	61.89	18.82	15.27						6.45		<b>0.34</b>
<b>#30</b>	13.1	1450	1500	Au	maj, cpx, st, CM		63.69	15.55	14.60						5.86		<b>0.10</b>
<b>#37</b>	15.3	1115	2940	Au	maj, cpx, st, mag, FeTi oxide		65.50	14.56	14.00	0.87		4.52					<b>0.77</b>
<b>#38</b>	15.3	1190	3150	Au	maj, cpx, st, mag, FeTi oxide		64.39	12.04	17.38	0.39		5.44					<b>0.23</b>
<b>#35</b>	15.3	1250	1920	Au	maj, cpx, st, CM	small melt pools	64.14	13.50	15.74						6.31		<b>0.13</b>
<b>#36</b>	20.7	1100	3240	Au	maj, st, cpx, mag, Na carb, FeTi oxide		64.76	12.43	15.27	2.00		3.17	2.38				<b>0.05</b>
<b>#27</b>	20.7	1200	1440	Au	maj, st, CM	Re furnace, small melt pools	76.30		17.56						5.50		<b>0.37</b>
<b>#28</b>	20.7	1300	1440	Au	maj, st, CM	Re furnace	75.32		17.71						5.88		<b>0.25</b>
<b>#19</b>	20.7	1400	3660	Au	maj, st, CM		75.14		18.42						5.81		<b>0.46</b>
<b>#11</b>	20.7	1450	1440	Au	maj, st, CM		76.15		17.66						5.59		<b>0.46</b>
<b>#13</b>	20.7	1480	1650	Au	maj, st, CM		76.75		17.03						5.89		<b>0.14</b>
<b>#16</b>	20.7	1530	1440	Au	maj, st, CM	small melt pools	76.48		17.41						5.99		<b>1.03</b>
<b>#31</b>	20.7	1600	600	Au	maj, st, CM		75.10		17.28						7.37		<b>0.34</b>
<b>#33</b>	20.7	1400	1440	Au	maj, st, CM	75% ATCM1 + 25% ATCM2	61.48		10.41						27.58		<b>2.16</b>

Mass balance calculations were performed as described in Supplementary Information. Phase proportions are in wt%. CM, carbonatite melt; cpx, clinopyroxene; cs, coesite; dol, dolomite; FeTi oxide, iron-titanium-rich oxide phase; gt, garnet; mag, magnesite; maj, majoritic garnet; Na carb, sodic carbonate; rut, rutile; SM, silicate melt; st, stishovite.

Extended Data Table 3 | Summary of reaction experiments run conditions and experimental products

Type	Starting Material	Expt No.	T (°C)	Duration (mins)	Mineral phases
Synthesis	KR4003 + 2.5 wt.% Fe “	G168	1600 <sup>‡</sup>	510	wad, maj, capv, Fe
		G176	1590	480	wad, maj, CM*
Reaction	G168 + ATCM2	G169	1600	30	maj, capv, rw, fper, CM, maj <sup>†</sup> , wad <sup>†</sup> , capv <sup>†</sup> , Di*
	G176 + ATCM2	G177	1590	124	maj, capv, rw, mag, CM, Al <sub>2</sub> O <sub>3</sub> , maj <sup>†</sup> , wad <sup>†</sup> , Di*
Mixture	<b>PM1:ATCM2 (wt.%) in Fe capsules</b>				
	9:1	Y17b	1400	180	fper, maj, rw, wad, capv*, melt, Fe, Di
	7:3	Y16a	1400	270	fper, maj, rw, wad, capv*, melt, Fe, Di
	1:1	Y16b	1400	270	fper, maj, rw, wad, capv*, melt, Fe, Di
	<b>PM1:ATCM2:Fe (mol.%) in an Au capsule</b>				
	16:4:1	G183b	1400		maj, rw, capv, mag, CM

capv, calcium perovskite; CM, carbonatite melt; Di, diamond; Fe, Fe metal; fper, ferropericlase; mag, magnesite; maj, majorite garnet; rw, ringwoodite; wad, wadsleyite.

\*Trace/minor phase.

<sup>†</sup>Relics of the peridotite starting material.

<sup>‡</sup>Thermocouple broke during run; temperature estimated using power curves with maximum uncertainty of  $\pm 150^\circ\text{C}$ .

# Holocene shifts in the assembly of plant and animal communities implicate human impacts

S. Kathleen Lyons<sup>1</sup>, Kathryn L. Amatangelo<sup>2</sup>, Anna K. Behrensmeyer<sup>1</sup>, Antoine Bercovici<sup>1</sup>, Jessica L. Blois<sup>3</sup>, Matt Davis<sup>1,4</sup>, William A. DiMichele<sup>1</sup>, Andrew Du<sup>5</sup>, Jussi T. Eronen<sup>6</sup>, J. Tyler Faith<sup>7</sup>, Gary R. Graves<sup>8,9</sup>, Nathan Jud<sup>10,11</sup>, Conrad Labandeira<sup>1,12,13</sup>, Cindy V. Looy<sup>14</sup>, Brian McGill<sup>15</sup>, Joshua H. Miller<sup>16</sup>, David Patterson<sup>5</sup>, Silvia Pineda-Munoz<sup>17</sup>, Richard Potts<sup>18</sup>, Brett Riddle<sup>19</sup>, Rebecca Terry<sup>20</sup>, Anikó Tóth<sup>1</sup>, Werner Ulrich<sup>21</sup>, Amelia Villaseñor<sup>5</sup>, Scott Wing<sup>1</sup>, Heidi Anderson<sup>22</sup>, John Anderson<sup>22</sup>, Donald Waller<sup>23</sup> & Nicholas J. Gotelli<sup>24</sup>

**Understanding how ecological communities are organized and how they change through time is critical to predicting the effects of climate change<sup>1</sup>. Recent work documenting the co-occurrence structure of modern communities found that most significant species pairs co-occur less frequently than would be expected by chance<sup>2,3</sup>. However, little is known about how co-occurrence structure changes through time. Here we evaluate changes in plant and animal community organization over geological time by quantifying the co-occurrence structure of 359,896 unique taxon pairs in 80 assemblages spanning the past 300 million years. Co-occurrences of most taxon pairs were statistically random, but a significant fraction were spatially aggregated or segregated. Aggregated pairs dominated from the Carboniferous period (307 million years ago) to the early Holocene epoch (11,700 years before present), when there was a pronounced shift to more segregated pairs, a trend that continues in modern assemblages. The shift began during the Holocene and coincided with increasing human population size<sup>4,5</sup> and the spread of agriculture in North America<sup>6,7</sup>. Before the shift, an average of 64% of significant pairs were aggregated; after the shift, the average dropped to 37%. The organization of modern and late Holocene plant and animal assemblages differs fundamentally from that of assemblages over the past 300 million years that predate the large-scale impacts of humans. Our results suggest that the rules governing the assembly of communities have recently been changed by human activity.**

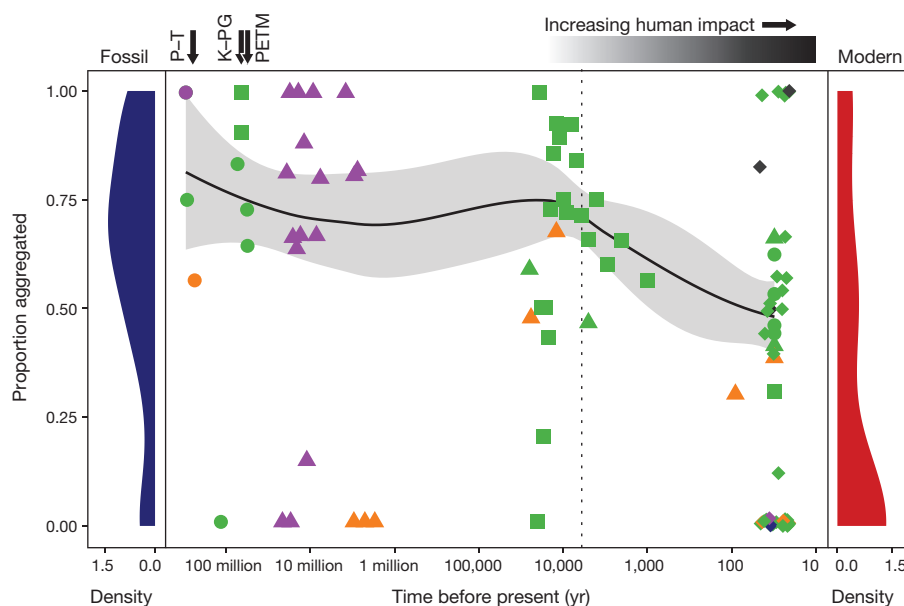
How are plant and animal communities organized, and does their structure change through time? This question has dominated many decades of research on community assembly rules and is critical to charting the future of biodiversity<sup>1</sup>. Whereas most studies have described overall community structure with simple indices such as species richness<sup>8</sup> and average co-occurrence<sup>3</sup>, some analyses categorize individual species pairs in assemblages as random, aggregated, or segregated<sup>2,9</sup>. Segregated species pairs may be generated by processes such as negative species interactions, distinct habitat preferences, and dispersal limitation. Aggregated species pairs may be generated by processes such as positive species interactions, shared habitat preferences, and concordant dispersal<sup>2</sup>. Recent meta-analyses document an

excess of segregated species pairs in modern communities: most significant species pairs co-occur less frequently than would be expected by chance<sup>2,10</sup>. The relative dominance of segregated versus aggregated species pairs suggests an important role for biotic interactions such as competition and predation, habitat selectivity, and dispersal limitation in structuring modern communities.

Do the patterns of species segregation that characterize modern assemblages also hold in the fossil record, or is the present different? If there was a change, when did the modern condition arise? There are many examples from the fossil record of times of major reorganization in ecological communities, such as a shift in the complexity of marine invertebrate communities after the end-Permian mass extinction<sup>11</sup>. But even during the lengthy periods between mass extinctions, the nature of species interactions may change. For example, the diversity and intensity of insect herbivory increased during a warming trend from the Late Palaeocene to the Eocene<sup>12</sup>. Moreover, many late Pleistocene plant and animal assemblages that contain some extant species have no modern analogues<sup>13,14</sup>. Such results hint that general patterns of species associations observed in contemporary assemblages could have been quite different in the past.

Here we ask whether non-random species associations of plant and mammal assemblages over the past 300 million years (Myr) are dominated by segregated or aggregated species pairs. This novel analysis is designed to compare statistical patterns of taxon associations for fossil and modern data using a consistent set of methodologies. We analysed 80 well-sampled fossil and recent assemblages: 38 for mammals and 42 for plants (see Supplementary Information, Extended Data Fig. 1 and Extended Data Table 1). Each data set contained information on taxon presence and absence across multiple localities in a particular time period (Extended Data Fig. 1 and Extended Data Table 1). Ages of plant data sets range from 307 million years ago (Ma) to the present and are from North America and Africa. Mammal data sets range in age from 21.4 Ma to the present and are from North America, Eurasia, and Africa. We compared each data set to a 'null' assemblage generated by randomization, scored each taxon pair as random, aggregated, or segregated, and used an empirical Bayes approach to control for the rate of false positive discoveries<sup>15</sup>; see Methods). Finally we

<sup>1</sup>Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington DC 20013, USA. <sup>2</sup>Department of Environmental Science and Biology, The College at Brockport – SUNY, Brockport, New York 14420, USA. <sup>3</sup>School of Natural Sciences, University of California, Merced, 5200 North Lake Road, Merced, California 95343, USA. <sup>4</sup>Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06520, USA. <sup>5</sup>Hominid Paleobiology Doctoral Program, Center for the Advanced Study of Hominid Paleobiology, Department of Anthropology, George Washington University, Washington DC 20052, USA. <sup>6</sup>Department of Geosciences and Geography, University of Helsinki, PO Box 64, 00014 University of Helsinki, Finland. <sup>7</sup>School of Social Science, The University of Queensland, Brisbane, Queensland 4072, Australia. <sup>8</sup>Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington DC 20013, USA. <sup>9</sup>Center for Macroecology, Evolution and Climate, University of Copenhagen, Copenhagen 2100, Denmark. <sup>10</sup>Biological Sciences Graduate Program, University of Maryland, College Park, Maryland 20742, USA. <sup>11</sup>Florida Museum of Natural History, University of Florida, Gainesville, Florida 32611, USA. <sup>12</sup>Department of Entomology, University of Maryland College Park, College Park, Maryland 20742, USA. <sup>13</sup>Key Lab of Insect Evolution and Environmental Changes, Capital Normal University, Beijing 100048, China. <sup>14</sup>Department of Integrative Biology and Museum of Paleontology, University of California Berkeley, Berkeley, California 94720, USA. <sup>15</sup>School Biology and Ecology & Sustainability Solutions Initiative, University of Maine, Orono, Maine 04469, USA. <sup>16</sup>Department of Geology, University of Cincinnati, Cincinnati, Ohio 45221, USA. <sup>17</sup>Department of Biological Sciences, Macquarie University, Sydney, New South Wales 2109, Australia. <sup>18</sup>Department of Anthropology, Human Origins Program, National Museum of Natural History, Smithsonian Institution, Washington DC 20013, USA. <sup>19</sup>School of Life Sciences, University of Nevada-Las Vegas, Las Vegas, Nevada 89154, USA. <sup>20</sup>Department of Integrative Biology, Oregon State University, Corvallis, Oregon 97331, USA. <sup>21</sup>Chair of Ecology and Biogeography, Nicolaus Copernicus University, Lwowska 1, 87-100 Toruń, Poland. <sup>22</sup>Evolutionary Studies Institute, University of the Witwatersrand, Jorissen Street, Braamfontein, Johannesburg 2001, South Africa. <sup>23</sup>Department of Botany, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA. <sup>24</sup>Department of Biology, University of Vermont, Burlington, Vermont 05405, USA.



**Figure 1 | Proportion of aggregated pairs over the past 300 Myr.**

Weighted Loess curve with shaded 95% confidence intervals illustrates reduction in the proportion of aggregated species pairs in the Holocene (log scale). Dotted vertical line at 5,998 years delineates the linear model breakpoint in the trend (Methods and Extended Data Fig. 2). Non-random species pairs of 'Fossil' data (blue density profile) are predominantly aggregated, whereas 'Modern' data (red density profile) are predominantly segregated. Colours indicate continent: North America (green), Eurasia (purple), Australia (dark grey), South America (dark blue), Africa

(orange). Point shapes indicate type of data: pollen (square), mammals (triangle), macroplants (circle). Data on terrestrial communities from ref. 2 are diamonds. All fossil and modern data are from mainland sites; no island sites were included. Time values of modern data points were assigned a single age (see Supplementary Information data sets), but are jittered for visual representation. P-T, Permo-Triassic transition; K-Pg, Cretaceous-Palaeogene transition; PETM, Palaeocene-Eocene thermal maximum.

synthesized our results with those from a meta-analysis of 39 modern communities that used the same methodology<sup>2,10</sup>.

For all fossil data sets, most taxon pairs were random (87–100% of possible pairs; Extended Data Table 1), which is also typical for modern assemblages<sup>2</sup>. This result reflects the statistically conservative nature of the tests used to identify significantly associated pairs, and the fact that most taxon pairs in a diverse, well-sampled assemblage interact weakly, or not at all. In 62 of 80 assemblages analysed here, a subset of taxon pairs showed significant associations that are stronger than can be explained by the null model, even after controlling for the false discovery rate (Fig. 1). Unlike modern mainland assemblages, most significant associations in the fossil record are aggregated, positive associations (Fig. 1). This pattern is consistent across the past 300 Myr for the diverse fossil assemblages in this study, which encompass mammals, plant macrofossils, and pollen from multiple continents and time slices.

However, beginning in the Holocene, there was a significant temporal trend towards a greater proportion of segregated species pairs, which is consistent with the results for modern assemblages. A breakpoint analysis indicates that the shift began approximately 6,000 years ago (Extended Data Fig. 2). Confidence intervals of the breakpoint are large owing to a lack of appropriate data sets between 20,000 and 1 million years ago. Therefore, it is difficult to pinpoint the exact time of the shift, but a closer examination of the data suggests that placing it within the Holocene is reasonable. Before the breakpoint, on average 64% of significant pairs were aggregated (median = 73%). After the breakpoint, the average dropped to 37% (median = 42%). This trend is not driven by the modern data and persists when only fossil data are analysed (Extended Data Fig. 3).

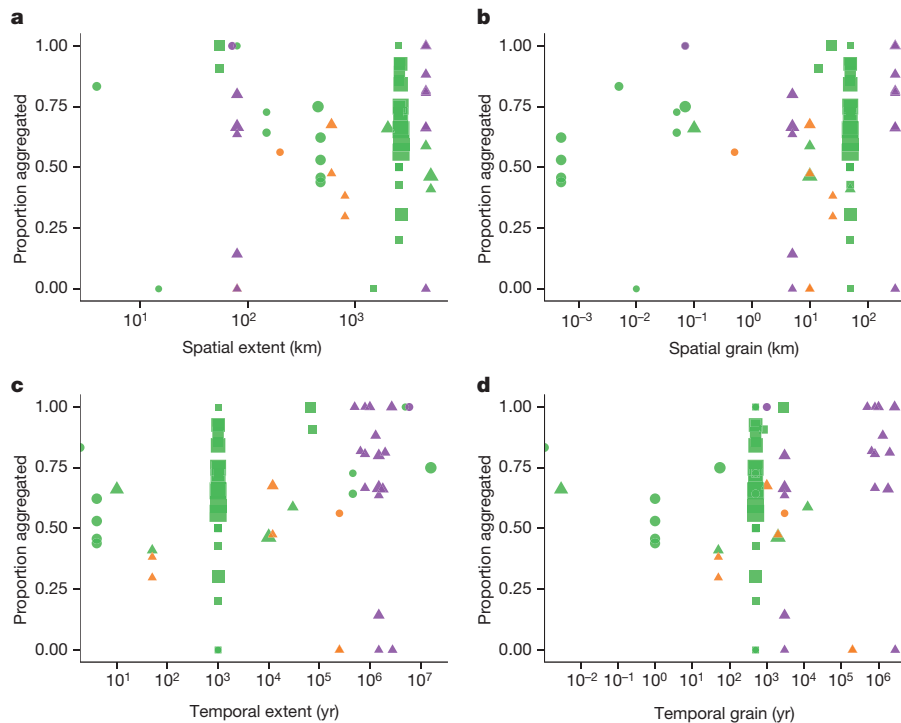
Why are species associations so different in fossil versus modern assemblages? We first tested and eliminated five potential 'artefact' hypotheses that are related to sampling issues (see Methods for details). (1) Collection modes were discounted because they were heterogeneous both for the modern and for fossil assemblages, and because the decrease in aggregated pairs was strong in fossil pollen

and mammal assemblages that spanned the shift. Moreover, sampling methodology was consistent within an assemblage type across periods that encompass the change (Extended Data Fig. 4). (2) Scale was discounted because there was no relationship between the spatial or temporal extent and grain of each data set and the percentage of aggregated pairs (Fig. 2 and Extended Data Fig. 5). (3) Taphonomic bias was discounted because the null model algorithm preserved the marginal totals of the data matrix in each randomized assemblage, controlling for simple taphonomic biases that could generate heterogeneity in the number of species per site or the number of occupied sites per species. (4) Taxonomic resolution was discounted because parallel analyses at the genus and species levels did not produce systematic changes in the proportions of aggregated pairs (Extended Data Table 2). (5) Increased sampling of rare species in modern data sets was discounted because segregated pairs tend to form in species with intermediate occupancy, whereas aggregated pairs form both in common and in rare species in modern and fossil data sets. All of these mechanisms can potentially affect assemblage structure in fossil and modern data sets. However, our analyses suggest that these mechanisms cannot account for the prominent decrease in aggregated species pairs that began during the Holocene (Fig. 1).

The failure of sampling issues to account for the temporal change in the percentage of non-randomly associated taxon pairs suggests that a mechanistic explanation is required. We consider two hypotheses that invoke a systematic change in either abiotic or biotic factors as drivers of co-occurrence patterns.

One of the most obvious differences between the present interval and the past 300 Myr of geological history represented by these fossil assemblages is the increasing variability of climate towards the present, associated with the glacial-interglacial cycles of the Quaternary period<sup>16</sup>. This is not to say that there were no periods of high climate variability before the ice ages, but that our data do not regularly sample times of high climate variability in deep time. If climate variability is responsible for the shift in the frequency of aggregated species pairs, there should be a negative relationship between climate variability





**Figure 2 | Relationship between scale and proportion of aggregated pairs.** The proportion of significant pairs that are aggregated does not depend on the temporal or spatial scale of data. Each point represents a single data set. **a, b**, Aggregated pairs versus spatial extent (longest linear distance between any two sites in a data set; **a**) or spatial grain (estimated radius of collection area that fossil specimens would have been transported

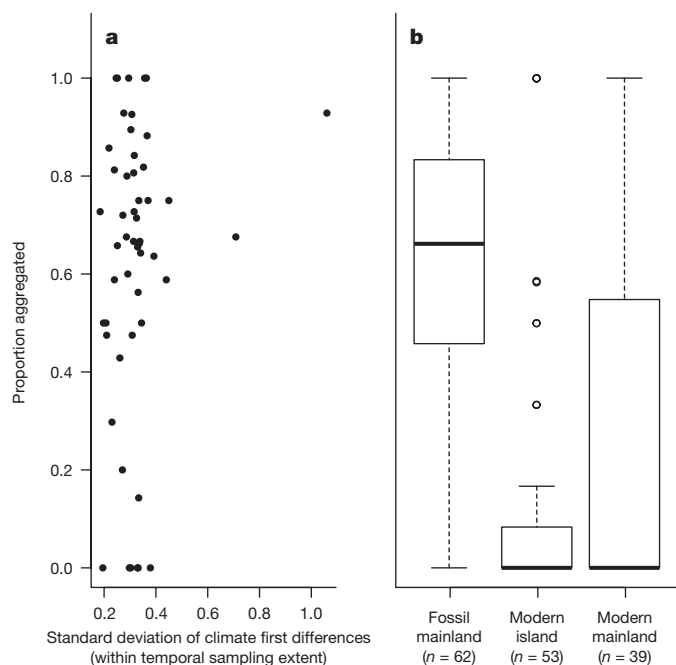
to the depositional environment in a typical locality; **b**). **c, d**, Proportion of aggregated pairs versus temporal extent (duration from the oldest to youngest locality in a data set; **c**) or temporal grain (typical amount of time-averaging of localities in a data set; **d**). Colours and shapes as in Fig. 1. Note the logarithmic scale of the *x* axes. Modern data from ref. 2 are excluded from this analysis.

and the percentage of aggregations. We quantified climate variability within the temporal extent of each data set for the past 65 Myr, using climate data from ice<sup>17</sup> and deep sea<sup>16</sup> cores that were standardized to a common scale (Methods). We found no relationship between the proportion of aggregated pairs and the standard deviation of climate within the sampled time slice (Extended Data Figs 6 and 7), or the standard deviation of the first differences of climate within the sampled time slice (Fig. 3a). Collectively, these results suggest that the increasing variability in climate in the Quaternary cannot explain the decreased frequency of aggregation.

An alternative explanation is that the mid- to late Holocene is unusual in the evolutionary history of terrestrial ecosystems, and that biotic drivers (as opposed to climate) now are different from what they have typically been over the past 300 Myr. First, we asked whether there was a significant shift in the proportions of aggregated versus segregated pairs across critical geological intervals that spanned periods of mass extinctions or major climatic change during the past 300 Myr (Extended Data Table 3 and Extended Data Fig. 4). We found a significant decrease in the percentage of positive associations only in data sets that spanned the Pleistocene–Holocene transition (11,700 years ago). With the exception of large-bodied mammals in North America and Africa (Extended Data Fig. 4b), aggregated species pairs decreased in all data sets through the Pleistocene–Holocene transition. In contrast, there was no significant change in the percentage of aggregations across the three other critical intervals that were encompassed by these data: the Permo–Triassic transition (252 Ma), the Cretaceous–Palaeogene transition (66 Ma), and the Palaeocene–Eocene thermal maximum (56 Ma). These intervals include the Cretaceous–Palaeogene mass extinction, responsible for the loss of the non-avian dinosaurs<sup>18</sup>, and the Permo–Triassic extinction, the largest mass extinction ever recorded<sup>18</sup>. Even the Palaeocene–Eocene thermal maximum, a period of major climatic change in which global temperatures increased  $\sim 5\text{--}8^\circ\text{C}$  in a few millennia<sup>19</sup>, did not coincide with a change in the relative proportions of aggregated versus segregated pairs.

It is difficult to pinpoint the exact mechanism responsible for the uniqueness of the present time interval. However, our analyses provide some clues about possible cause. Data that encompass the shift towards the modern pattern are almost exclusively North American (Fig. 1 and Extended Data Fig. 4). The statistical confidence interval bracketing the breakpoint at 6,000 years ago encompasses the beginning of agriculture in North America around 8,000 years ago<sup>6</sup> and the increase in human populations during the Holocene<sup>4,5</sup>. The trend towards greater segregations in North American pollen (Fig. 1 and Extended Data Fig. 4), with particularly strong shifts occurring in the past 2,000 years<sup>20</sup>, is also consistent with the history of agriculture in North America. Cultivation of multiple species of domesticated plants began approximately 3,800 years ago<sup>6,7</sup>, with evidence for more general dependency on agriculture in North America beginning 1,300 years ago<sup>6,7</sup>. Estimates of global land area under cultivation increase rapidly starting 6,000 years ago and are as high as  $4 \times 10^8$  hectares (1 hectare =  $10^4 \text{ m}^2$ ) by 2,000 years ago<sup>4,5</sup>.

Possible drivers by which increasing human impacts led to an increase in segregated pairs include (1) increases in hunting and domestication of particular species<sup>21,22</sup>, (2) changes in land use<sup>4,5</sup>, (3) increases in the frequency of fire<sup>4</sup>, (4) increases in habitat fragmentation and dispersal barriers<sup>23,24</sup>, and (5) deliberate and accidental spread of species beyond their native geographical ranges<sup>25–27</sup>. We note that modern island assemblages (which we excluded from our comparisons with fossil assemblages) are more segregated than modern mainland assemblages (Fig. 3b), which is consistent with the hypothesis that habitat fragmentation and dispersal limitation have increased segregated pairs. Possibly all of the processes listed play a role. Although their combined effects on taxon pairs are difficult to predict, the relative importance of factors structuring species co-occurrence appears to have changed through the Holocene. Future work comparing the co-occurrence structure of fossil and modern communities should allow us to better understand how this alteration will play out in the future. Regardless of the precise mechanisms,



**Figure 3 | Tests of possible mechanisms for decreasing spatial aggregation through time.** **a**, Climate variability within the temporal extents of the fossil data sets is uncorrelated with the proportion of aggregated species pairs. Climate variability is measured as the standard deviation of the first differences in climate across the interval (see also Extended Data Fig. 6). **b**, Box plots show the proportion of aggregated species pairs for fossil data, modern islands and modern mainland assemblages. Dashed lines indicate maximum and minimum values, circles are outliers. Island assemblages, with more limited capacity for dispersal, have the smallest and least variable fraction of aggregated pairs. Mainland fossil assemblages are significantly more aggregated than mainland modern assemblages. Note, these island data were excluded from other analyses.

humans appear to be agents of disturbance on a large scale and have been so for longer than is often recognized.

Our results suggest that assemblage co-occurrence patterns remained relatively consistent for 300 Myr but have changed over the Holocene as the impact of humans has dramatically increased. Across shallower time intervals, other studies have documented that local and regional species composition has changed substantially over recent decades<sup>28,29</sup> and millennia<sup>30</sup>. The rules governing community assembly, at least as implicated by co-occurrence patterns, seem to have changed during the Holocene and continue to change with the increasing influence of human activity. The co-occurrence structure of modern and recent plant and animal assemblages thus appears to be unique in the evolutionary history of terrestrial ecosystems, an important perspective for assessing challenges to these ecosystems in the face of present and future human impacts.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 August; accepted 13 November 2015.

Published online 16 December 2015; corrected online 6 January 2016 (see full-text HTML version for details).

1. Botkin, D. B. *et al.* Forecasting the effects of global warming on biodiversity. *Bioscience* **57**, 227–236 (2007).
2. Gotelli, N. J. & Ulrich, W. The empirical Bayes approach as a tool to identify non-random species associations. *Oecologia* **162**, 463–477 (2010).
3. Gotelli, N. J. & McCabe, D. J. Species co-occurrence: a meta-analysis of J.M. Diamond's assembly rules model. *Ecology* **83**, 2091–2096 (2002).
4. Marlon, J. R. *et al.* Global biomass burning: a synthesis and review of Holocene paleofire records and their controls. *Quat. Sci. Rev.* **65**, 5–25 (2013).
5. Puddiman, W. F. & Ellis, E. C. Effect of per-capita land use changes on Holocene forest clearance and CO<sub>2</sub> emissions. *Quat. Sci. Rev.* **28**, 3011–3015 (2009).

6. Zeder, M. A. & Smith, B. D. A conversation on agricultural origins: talking past each other in a crowded room. *Curr. Anthropol.* **50**, 681–691 (2009).
7. Smith, B. D. & Yarnell, R. A. Initial formation of an indigenous crop complex in eastern North America at 3800 B.P. *Proc. Natl Acad. Sci. USA* **106**, 6561–6566 (2009).
8. Connor, E. F. & McCoy, E. D. The statistics and biology of the species-area relationship. *Am. Nat.* **113**, 791–833 (1979).
9. Connor, E. F., Collins, M. D. & Simberloff, D. The checkered history of checkerboard distributions. *Ecology* **94**, 2403–2414 (2013).
10. Ulrich, W. & Gotelli, N. J. Null model analysis of species associations using abundance data. *Ecology* **91**, 3384–3397 (2010).
11. Wagner, P. J., Kosnik, M. A. & Lidgard, S. Abundance distributions imply elevated complexity of post-Paleozoic marine ecosystems. *Science* **314**, 1289–1292 (2006).
12. Wilf, P. & Labandeira, C. C. Response of plant-insect associations to Paleocene-Eocene warming. *Science* **284**, 2153–2156 (1999).
13. Lyons, S. K. A quantitative model for assessing community dynamics of Pleistocene mammals. *Am. Nat.* **165**, E168–E185 (2005).
14. Jackson, S. T. & Williams, J. W. Modern analogs in Quaternary paleoecology: here today, gone yesterday, gone tomorrow? *Annu. Rev. Earth Planet. Sci.* **32**, 495–537 (2004).
15. Ulrich, W. *Pairs* – a FORTRAN program for studying pair wise species associations in ecological matrices. <http://www.keib.umk.pl/wp-content/uploads/2013/04/PairsManual.pdf> (2008).
16. Zachos, J. C., Dickens, G. R. & Zeebe, R. E. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* **451**, 279–283 (2008).
17. Augustin, L. *et al.* EPICA community members. Eight glacial cycles from an Antarctic ice core. *Nature* **429**, 623–628 (2004).
18. Bambach, R. K. Phanerozoic biodiversity mass extinctions. *Annu. Rev. Earth Planet. Sci.* **34**, 127–155 (2006).
19. Bowen, G. J. *et al.* Two massive, rapid releases of carbon during the onset of the Palaeocene-Eocene thermal maximum. *Nature Geosci.* **8**, 44–47 (2015).
20. Blois, J. L. *et al.* A framework for evaluating the influence of climate, dispersal limitation, and biotic interactions using fossil pollen associations across the late Quaternary. *Ecography* **37**, 1095–1108 (2014).
21. Delcourt, P. A. & Delcourt, H. R. *Prehistoric Native Americans and Ecological Change: Human Ecosystems in Eastern North America since the Pleistocene*. (Cambridge Univ. Press, 2004).
22. Barnosky, A. D. Colloquium paper: megafauna biomass tradeoff as a driver of Quaternary and future extinctions. *Proc. Natl Acad. Sci. USA* **105** (suppl. 1), 11543–11548 (2008).
23. Haddad, N. M. *et al.* Habitat fragmentation and its lasting impact on Earth's ecosystems. *Sci. Adv.* **1**, e1500052 (2015).
24. Williams, M. *Deforesting the Earth: From Prehistory To Global Crisis* (Univ. Chicago Press, 2003).
25. Grayson, D. K. The archaeological record of human impacts on animal populations. *J. World Prehist.* **15**, 1–68 (2001).
26. van Kleunen, M. *et al.* Global exchange and accumulation of non-native plants. *Nature* **525**, 100–103 (2015).
27. Capinha, C., Essl, F., Seebens, H., Moser, D. & Pereira, H. M. The dispersal of alien species redefines biogeography in the Anthropocene. *Science* **348**, 1248–1251 (2015).
28. Dornelas, M. *et al.* Assemblage time series reveal biodiversity change but not systematic loss. *Science* **344**, 296–299 (2014).
29. Tóth, A. B., Lyons, S. K. & Behrensmeier, A. K. A century of change in Kenya's mammal communities: increased richness and decreased uniqueness in six protected areas. *PLoS ONE* **9**, e93092 (2014).
30. Thompson, J. R., Carpenter, D. N., Cogbill, C. V. & Foster, D. R. Four centuries of change in northeastern United States forests. *PLoS ONE* **8**, e72540 (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank G. Dietl for comments that improved the manuscript. Support for this research was provided by a National Museum of Natural History Program grant to the Evolution of Terrestrial Ecosystems Program (ETE) and NSF-DEB 1257625. This is ETE publication 338.

**Author Contributions** All authors provided input into the final manuscript. S.K.L. helped design the study, contributed and analysed data, and wrote the paper. K.L.A., J.L.B., J.T.F., and W.U. were involved in the study design, and contributed and analysed data. A.T. was involved in the study design, contributed and analysed data, and created Supplementary Fig. 1. A.K.B., A.B., W.A.D., J.T.E., N.J., C.L., C.V.L., S.W., and B.R. were involved in the study design and contributed data. H.A., J.A., and D.W. contributed data. M.D. and J.H.M. were involved in the study design, analysed data, created figures and wrote text for the supplementary information. A.D. was involved in the study design and analysed data. G.R.G., B.M., D.P., S.P.-M., R.T., R.P., and A.V. were involved in the study design. N.J.G. designed the study and helped write the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.K.L. (lyonss2@si.edu).

## METHODS

No statistical methods were used to predetermine sample size.

**Detection of non-random species pairs.** The data for each analysis consist of a binary presence–absence matrix in which each row is a taxon and each column is a sample. The entries represent the presence (1) or absence (0) of a particular taxon in a particular sample. Within this matrix, each of the  $S(S-1)/2$  unique species pairs is tested and classified as random, aggregated, or segregated. The tests were performed with the PAIRS version 1.0 software application<sup>15,31</sup>. The methodology is described fully in ref. 2 and is briefly described here.

The analysis begins by calculating a scaled  $C$  score<sup>32</sup>:  $C_{ij} = (R_i - D)(R_j - D)/R_i R_j$ , where  $C_{ij}$  is the  $C$  score for species pair  $i$  and  $j$ ,  $R_i$  is the row total (the number of species occurrences) for species  $i$ ,  $R_j$  is the row total for species  $j$ , and  $D$  is the number of shared sites in which both species are present. For each species pair, this index ranges from 0.0 (aggregation: maximal co-occurrence of both species) to 1.0 (segregation: minimal co-occurrence of both species). PAIRS calculates the  $C$  score for each pair of species and assigns it to a histogram bin. There are 20 bins that range from 0 to 1 in 0.05 intervals, plus a bin for 0.0 (perfectly aggregated pairs) and a bin for 1.0 (perfectly segregated pairs).

We next estimate the  $P$  value associated with each species pair by a randomization test. The data matrix is first randomized by reshuffling all matrix elements, with the restriction that the row and column sums of the original matrix are preserved. This ‘fixed-fixed’ algorithm has been subject to extensive benchmark testing with artificial random and structured matrices<sup>2,33,34</sup>. Compared with a variety of other null model algorithms, the fixed-fixed algorithm most effectively screens against type I errors (incorrectly rejecting the null hypothesis for a random matrix), but is somewhat conservative<sup>33</sup>.

An alternative algorithm ‘fixed-equiprobable’ retains row sums (species occurrence frequencies), but allows column totals (species richness per site) to vary freely. The fixed-equiprobable algorithm also has good statistical properties, and is appropriate for modern data sets in which sampling effort has been standardized, such as quadrat samples of fixed area. However, this algorithm is not appropriate for fossil data because the number of species detected per site in fossil assemblages is determined primarily by sampling effort of the collector and by site-specific taphonomic biases in preservation.

For these reasons, we have used only the fixed-fixed model, both for the analysis of fossil assemblages and for comparison with modern assemblages. Details of the randomization are discussed further in refs 2, 35. Using 1,000 randomizations, we create a simple  $P$  value (two-tailed test) for each species pair, which leads to a classification of each species pair as aggregated, random, or segregated.

However, with a total of  $S(S-1)/2$  unique pairs in a matrix of  $S$  species, retaining all of the significant pairs ( $P < 0.05$ ) would generate a potentially large number of false positive results. This problem has frequently arisen in the analysis of micro-arrays, genomic surveys, and other examples of ‘big data’<sup>36</sup>. The PAIRS analysis relies on an empirical Bayes approach by creating a histogram of  $C$  score values based on the pairs generated in each null assemblage. To screen out false positives, we calculated the average number of species pairs in each bin of the histogram. Next, we determined the observed number of pairs from the empirical assemblage in each bin, ordered by their  $P$  values from the simulation. We retained only those pairs that were above the mean number for each bin and were statistically significant on the basis of the simple  $P$  value criterion. This double screen effectively eliminates many of the false-positives that can arise in random data sets<sup>2</sup>.

**Weighted Loess regression.** A Loess smoothing line was created with the `stat_smooth` function in the R package `ggplot2` version 1.0.0 (ref. 37) using default parameters. For Loess fitting, the fit at point  $x$  is made using points in the neighbourhood of  $x$  (closest 75% of total points), with tricubic weighting (proportional to  $(1 - (\text{distance}/\text{maximum distance}))^3$ ). Points were additionally weighted by the number of sites in each matrix and 95% confidence intervals were generated using a  $t$ -based approximation.

**Analysis of climate variability.** To examine the how climate variability impacts the percentage of aggregated species pairs, we used climate proxy data from ice<sup>17</sup> and deep sea cores<sup>16</sup>, which collectively encompass the past 65 Myr of the assembled data sets. The European Project for Ice Coring in Antarctica (EPICA) data were used preferentially when there was temporal overlap between proxy data sets. Climate data were mean centred and standardized before pooling into a single data time series. We then sampled the climate data across the ‘temporal extents’ (Extended Data Table 1) of the individual Evolution of Terrestrial Ecosystems Program (ETE) data sets to test if there were relationships between the percentage of aggregated species pairs and climate variability. Climate variability was calculated in two ways: (1) as the standard deviation of climate within the temporal extent of each data set and (2) as the standard deviation of the first differences (changes in climate from available time-step to time-step within the temporal extent of a data set) of climate. We used standard deviation because it helps address

issues with changes in data density over time. Estimated rates of change are sensitive to the time span over which they are measured and more closely spaced data would shift apparent rates of change. Approaches using standard deviation are less sensitive to this issue. We also compared climate variability with age (years before present) of ETE data sets to test for Phanerozoic-scale trends in climate variability sampled by ETE data sets.

**Breakpoint analysis.** We used a maximum likelihood approach, available in the R package ‘segmented’ version 1.1, to estimate the breakpoint time at which the sharper decline in aggregated species pairs began. This analysis used an initial linear model of the proportion of aggregated pairs as a function of community age ( $\log_{10}$  of years before present) to generate a best-fitting number of breakpoints, with separate regression lines fit to each segment. A bootstrap of 1,000 replicates was used to estimate uncertainty in the model parameters (including uncertainty in the time of the breakpoint).

**Tests of artefacts. Collection modes.** We thought that differences in the way fossil and modern data are collected might be responsible for the observed difference in the relative proportions of aggregated versus segregated species pairs in modern communities<sup>2,10</sup> and fossil communities. There are two reasons why collection modes are not likely to be responsible for this difference. First, fossil collections are heterogeneous by nature. Different collecting methods are used for different taxonomic groups (for example, bulk sampling, surface sampling, cores). Moreover, even within a taxonomic group, the type of depositional environment imposes different kinds of bias (for example, cave sites versus open pits for Pleistocene mammals). Second, we see a switch from species pairs that are dominated by aggregations to those dominated by segregations in our data sets that span the Pleistocene–Holocene transition (Extended Data Fig. 4 and Extended Data Table 1). In particular, mammal assemblages show a switch from  $>50\%$  aggregations in the Pleistocene to  $<50\%$  aggregations in the Holocene. The data encompassing this switch are all fossil localities and there are similar biases in both time slices. Although there is variation in the pollen assemblages, a weighted regression that takes into account the sampling in each time slice shows a significant decrease through time ( $P = 0.04$ ,  $R^2 = 0.15$ ). This trend of increasing percentage of segregated pairs begins approximately 14,000 years ago and continues across the Holocene with the switch occurring in the final 1,000 year time slice<sup>20</sup>. The fact that these data were all collected using the same sampling techniques suggests that sampling cannot account for this pattern.

**Issues of scale.** It is generally assumed that fossil data are biased. Although the type of bias is not the same for all taxonomic groups, most fossil assemblages contain some degree of temporal or spatial averaging<sup>38</sup>. That is, they represent accumulations of species that can occur over hundreds or thousands of years and may mix species that did not exist at the locality at the same time<sup>39</sup>. The fossil data sets in this analysis include assemblages that range from no time-averaging (for example, fossil leaves preserved in volcanic event beds) to those that are time-averaged over thousands or hundreds of thousands of years (for example, some mammal assemblages). In addition, some data sets could not be resolved to time bins of less than a million years. Spatial averaging is less of an issue in these data sets, but individual samples are drawn from areas with diameters ranging from a few metres to more than 300 km (Supplementary Table 1).

If issues of scale are contributing to the pattern found here, there should be a relationship between the proportion of significant pairs that are aggregated and the spatial or temporal scale of the data. We evaluated this by estimating the spatial or temporal grain and extent of each data set included in the analyses (Extended Data Table 1) and determining if there was a significant relationship with the percentage of aggregations. The spatial grain is the estimated radius of collection area over which fossil specimens would have been transported to the depositional environment in a typical locality. The temporal grain is the typical amount of time-averaging of localities in a data set. Spatial extent is the longest linear distance between any two sites in a data set and temporal extent is the duration from the oldest to youngest locality in a data set.

We found no relationship between the scale of the data sets and the proportion of significant pairs that were aggregated versus segregated (Fig. 2 and Extended Data Fig. 5). Regression analyses were not significant and explained very little of the variation in the data (Extended Data Fig. 6). The pattern of segregated versus aggregated pairs was not different in fossil versus modern assemblages because of biases related to the scale of fossil data.

**Taphonomic bias.** How can taphonomy and palaeoenvironment affect species frequencies (richness) and spatial representation? The fossil record contains buried assemblages of species that were derived from living communities at different times in the past. Species representation (presence or absence) in individual fossil assemblages is a critical attribute of our data sets, therefore we need to consider how this variable might be biased relative to original associations of species in communities. Taphonomic processes operate during the transition of dead remains into preserved samples and thus control the biological information that passes



from the living community into the fossil record<sup>39</sup>. These processes act as filters that can alter species representation in fossil samples in a variety of ways: (1) selective preservation of organisms with particular body compositions and sizes, for example organisms with and without mineralized skeletons, larger versus smaller individuals; (2) variable preservation of organisms depending on their population abundance, spatial distribution and life habits, for example aquatic versus terrestrial; (3) post-mortem or depositional mixing of species that did not live together (time-averaging), or separation of species that did (selective transport or destruction). Additionally, some types of environment are better represented in the depositional record than others, such as wetlands versus dry land surfaces. All of these add up to potential biases that could affect biological signals and the proportions of random versus significant species pairs, or the proportions of segregated versus aggregated pairs, in our analyses.

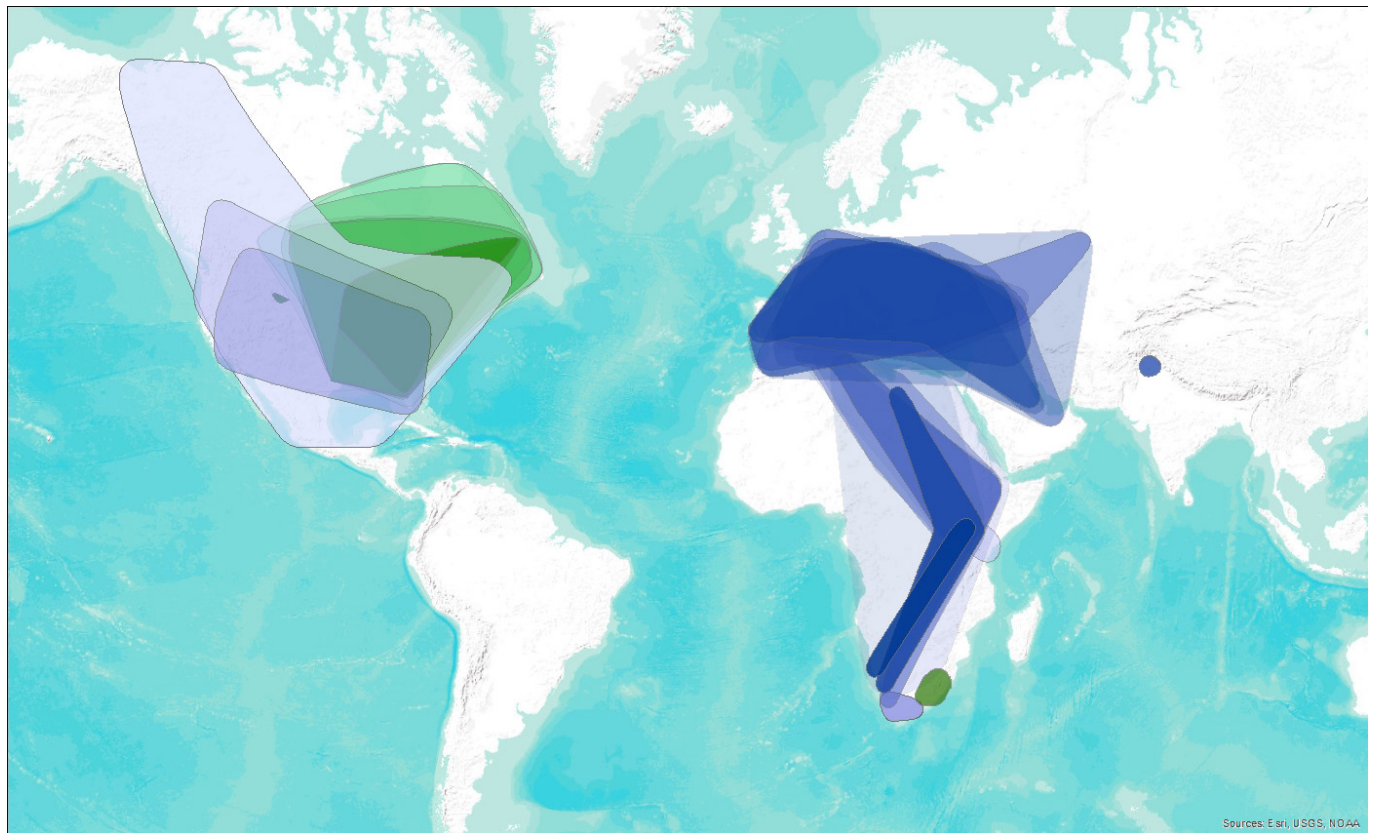
However, the particular null model algorithm used effectively controls for major sources of taphonomic bias in the data set. This 'fixed-fixed' algorithm<sup>33</sup> creates null assemblages that have the same species richness per sample, and the same number of occurrences per species, as the original data. Thus, if there are preservation biases that generate heterogeneity in the total number of fossil species per sample, or biases in the number of specimens per species, these are effectively controlled for in the analysis. Significant patterns of species aggregation are those measured beyond the effects of sampling heterogeneity in the occurrences of species or the number of species per sample. Similar sampling effects are controlled for in the modern data, which can also exhibit variation in the commonness or rarity of species and in the number of species per sample.

**Taxonomic resolution of the data.** Fossils are not always resolvable to the species level and are frequently analysed at the genus level. This may have the effect of increasing geographical ranges and overlap between taxa, and may contribute to the dominance of aggregated pairs found in this study. To test whether this was the case, we analysed 18 of the data sets at the species and genus level (16 mammal and 2 plant data sets). If taxonomic resolution is driving the pattern, we expect to see an increase in the proportion of aggregated pairs when species are lumped into genera. We found that six of the data sets showed the expected increase. However, nine showed a decrease and three showed no change (Extended Data Table 2). Interestingly, one of the modern data sets on small mammals from the Great Basin had genetic information that indicated that some were cryptic species. When the analysis was re-run with the cryptic species identified, there was an increase in the proportion of significantly aggregated pairs (from 50% to 61%). This is in the opposite direction that we would expect if lumping species into genera artificially increased aggregated pairs. Taken together, these results suggest that the differences between species associations over the past 300 Myr and the present are not driven by the taxonomic resolution of fossil data.

**Sampling of abundant and rare species in fossil and modern data.** The results of null model analyses of abundance versus presence-absence data are compared in ref. 10. The two kinds of analysis give qualitatively comparable results, although the abundance analyses are somewhat more powerful in detecting non-randomness. It is generally assumed that fossil deposits miss the rarest species in a community because preservation potential increases with abundance; more individuals means more opportunities for fossilization events. If rare species are more likely to form segregated pairs, we would expect to see more segregations in modern data sets because they should sample more of the rare species than comparable fossil data sets. Within fossil data sets, we would expect to see more segregated pairs in data sets with better sampling and more rare species. We evaluated this using a data visualization technique. We present the results of our analyses as a series of pairwise species by species matrices and order species by occupancy (see Supplementary Information: data sets). Occupancy decreases as one moves to the right on the *x* axis and up on the *y* axis. Species with the highest occupancy are close to the origin. The pairwise associations are colour-coded: grey for random pairs, blue for aggregated pairs, and red for segregated pairs. If increasing sampling of rare species is responsible for the pattern we document, then we would expect to see a preponderance of red, segregated pairs in the upper, right-hand corner of the species by species matrices. In particular, this should show up in data sets with better sampling and those that encompass the shift from more aggregated to more segregated pairs (for example, Pleistocene-Holocene mammals and pollen, modern mammals in Kenya, and modern plants in Wisconsin). This is not the pattern that we see. In fact, we find that segregated pairs tend to form with species of intermediate occupancy and that aggregated pairs form both with common species and with rare species. Differences in the sampling of rare species between fossil and modern data sets cannot account for the shift in species associations over time.

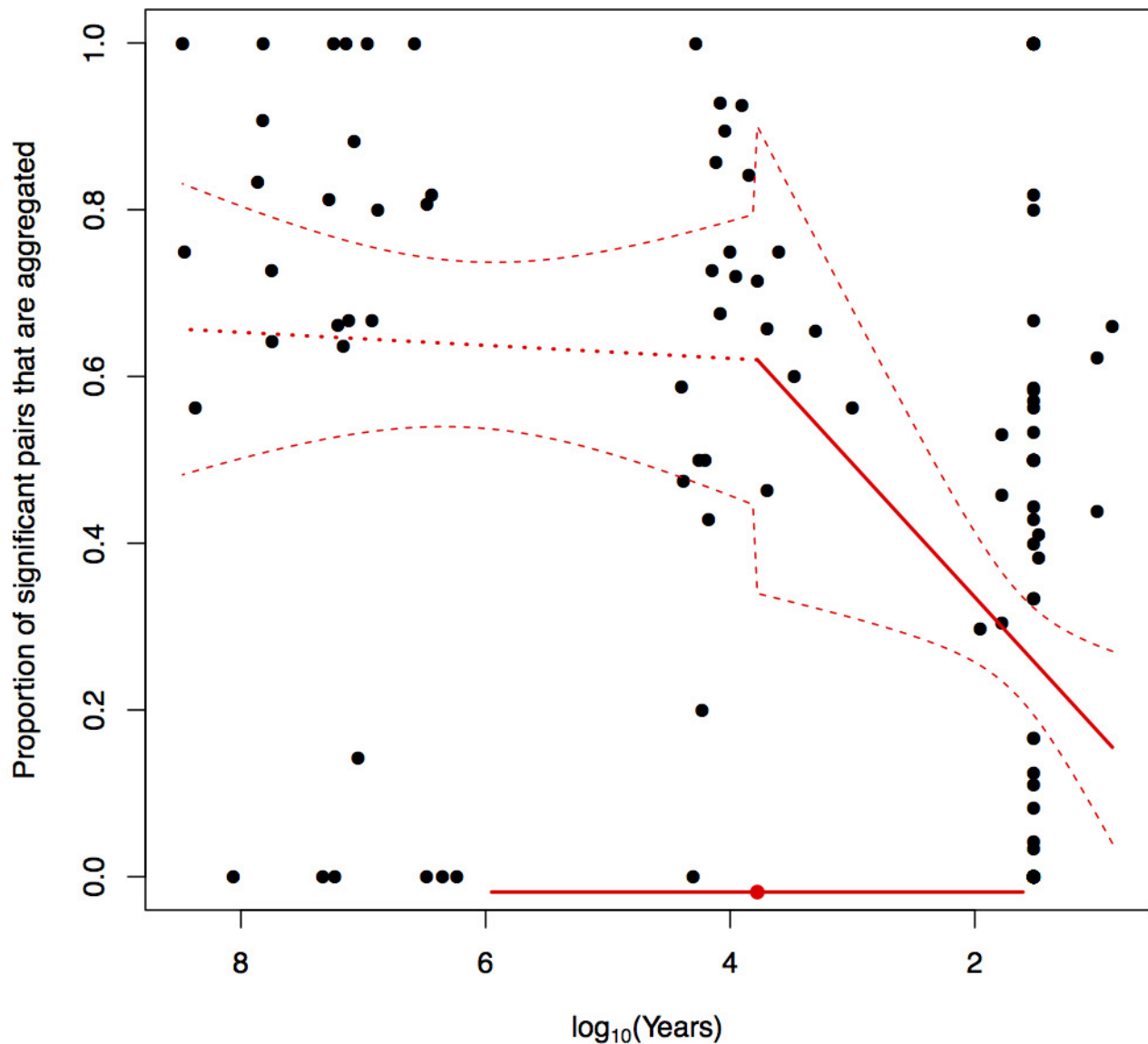
31. Ulrich, W. Pairs – a FORTRAN program for studying pair-wise species associations in ecological matrices. <http://www.keib.umk.pl/pairs/?lang=en>, (2010).
32. Stone, L. & Roberts, A. The checkerboard score and species distributions. *Oecologia* **85**, 74–79 (1990).
33. Gotelli, N. J. Null model analysis of species co-occurrence patterns. *Ecology* **81**, 2606–2621 (2000).
34. Ulrich, W. & Gotelli, N. J. Null model analysis of species nestedness patterns. *Ecology* **88**, 1824–1831 (2007).
35. Gotelli, N. J. & Ulrich, W. Statistical challenges in null model analysis. *Oikos* **121**, 171–180 (2012).
36. Efron, B. Bayesians, frequentists, and scientists. *J. Am. Stat. Assoc.* **100**, 1–5 (2005).
37. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009).
38. Kidwell, S. M. & Holland, S. M. The quality of the fossil record: implications for evolutionary analyses. *Annu. Rev. Ecol. Syst.* **33**, 561–588 (2002).
39. Behrensmeier, A. K., Kidwell, S. M. & Gastaldo, R. A. Taphonomy and paleobiology. *Paleobiology* **26**, 103–147 (2000).





**Extended Data Figure 1 | Map showing distribution of fossil data sets.** Polygons enclose the localities for each fossil data set included in our analyses. Mammals are in blue, plants are in green. Dark colours represent data sets that are older. This map was created using ArcGIS software by Esri and can be found at <http://www.arcgis.com/home/item.html?id=c61ad8ab017d49e1a82f580ee1298931>.

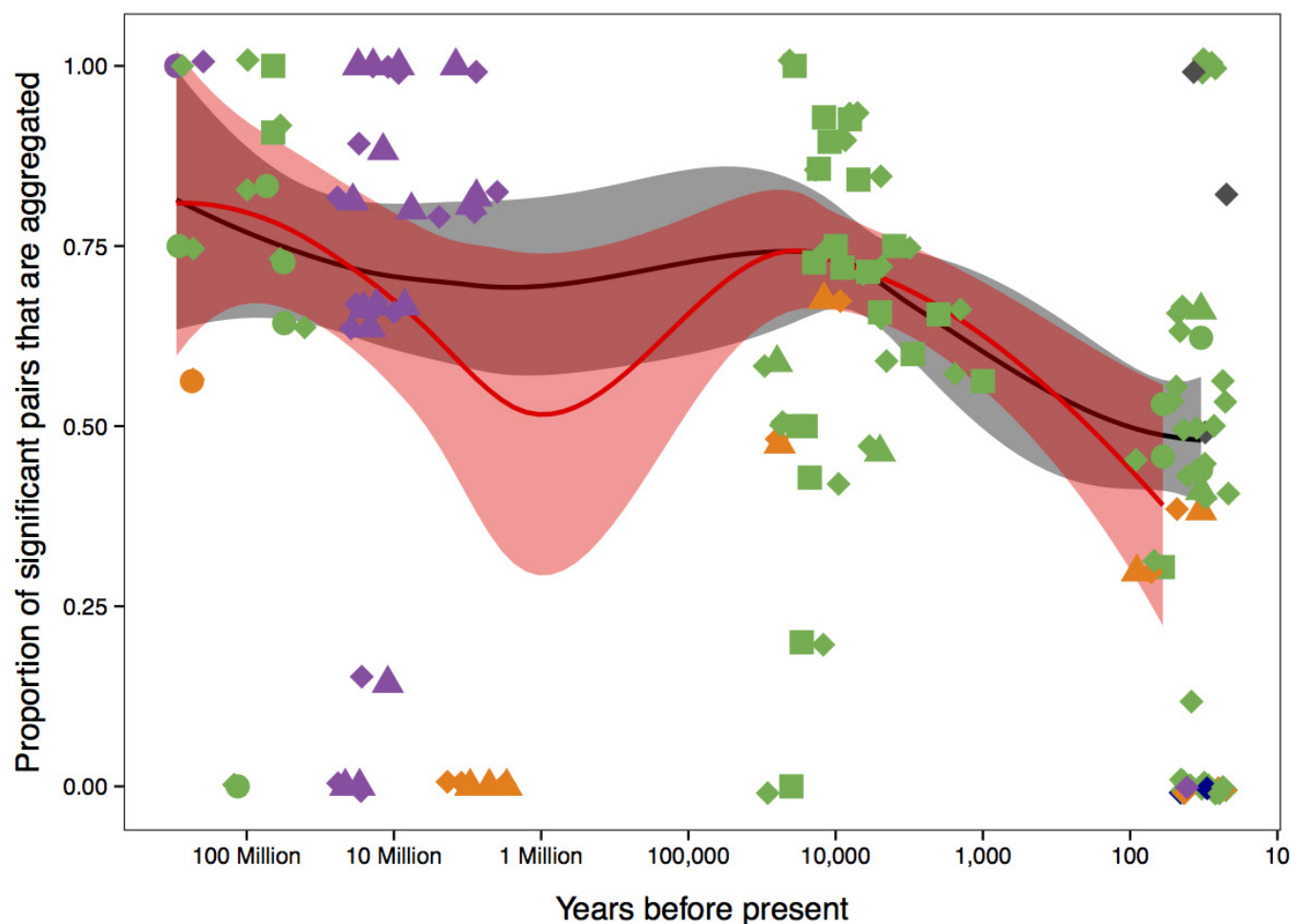
c61ad8ab017d49e1a82f580ee1298931. ArcGIS and ArcMap are the intellectual property of Esri and are used herein under license. Copyright © Esri. All rights reserved. For more information about Esri software, please visit <http://www.esri.com>.



#### Extended Data Figure 2 | Breakpoint analysis of the composite data.

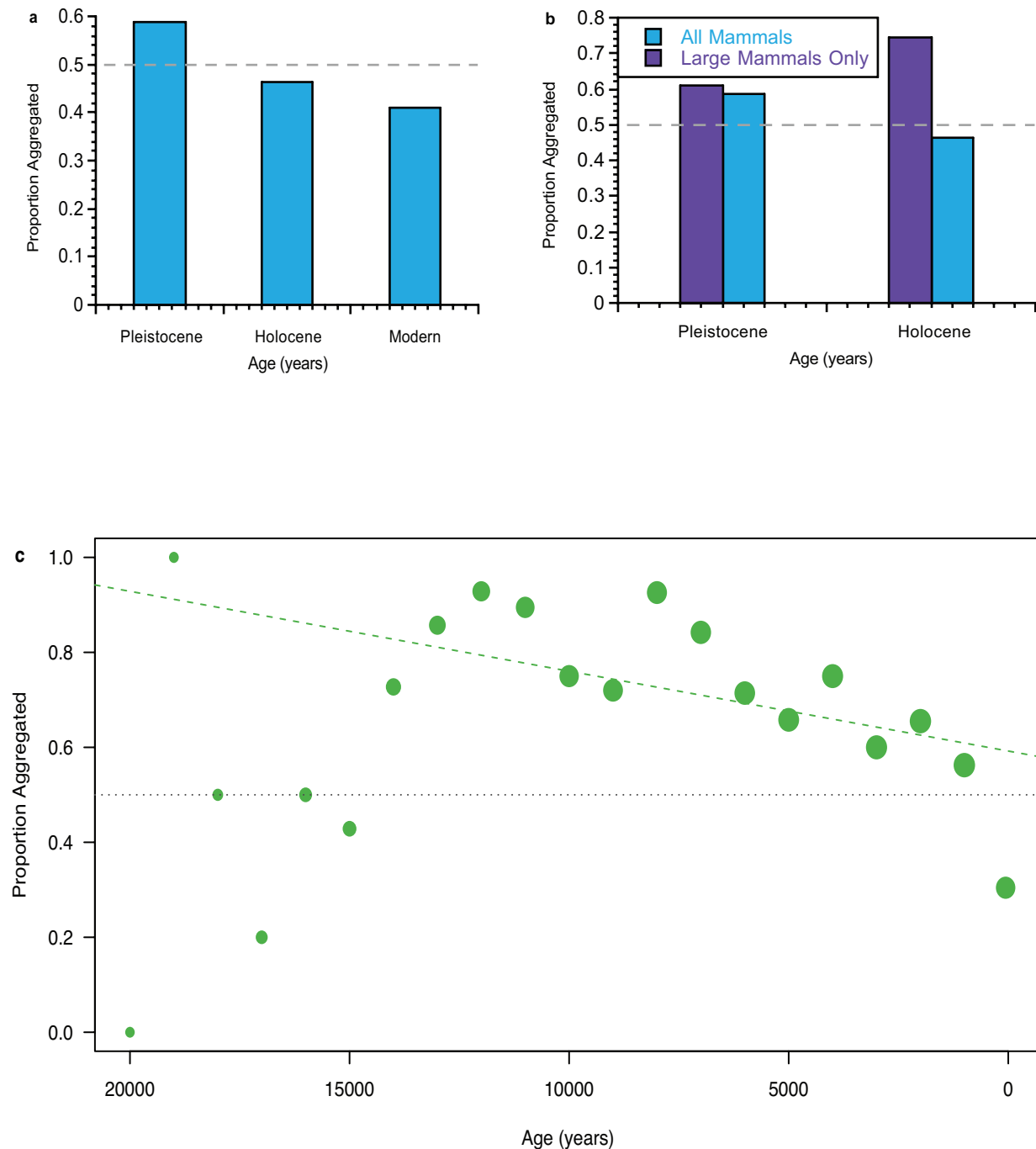
The analysis was performed on all data including the islands (see Fig. 3 main text), showing the mean estimate (red point;  $10^{3.778}$  years) and 95% confidence interval (error bar at base of plot;  $10^{1.606}$ ,  $10^{5.951}$  years) of the initiation of reduced percentage of aggregated species pairs, as well as

the mean and confidence intervals around the change in slope of the two resulting linear models. The breakpoint analysis was run using all the data resolved to the best possible dates to allow for the greatest amount of power in detecting where the switch occurred. However, the results were similar when island data were excluded.



**Extended Data Figure 3 | Weighted Loess curve with and without modern data.** Loess curve weighted by number of sites with shaded 95% confidence intervals illustrates the reduction in the proportion of aggregated species pairs towards the present. Data are analysed with (black line and shading) and without (red line and shading) the modern data. Colours indicate continent: North America (green), Eurasia (purple),

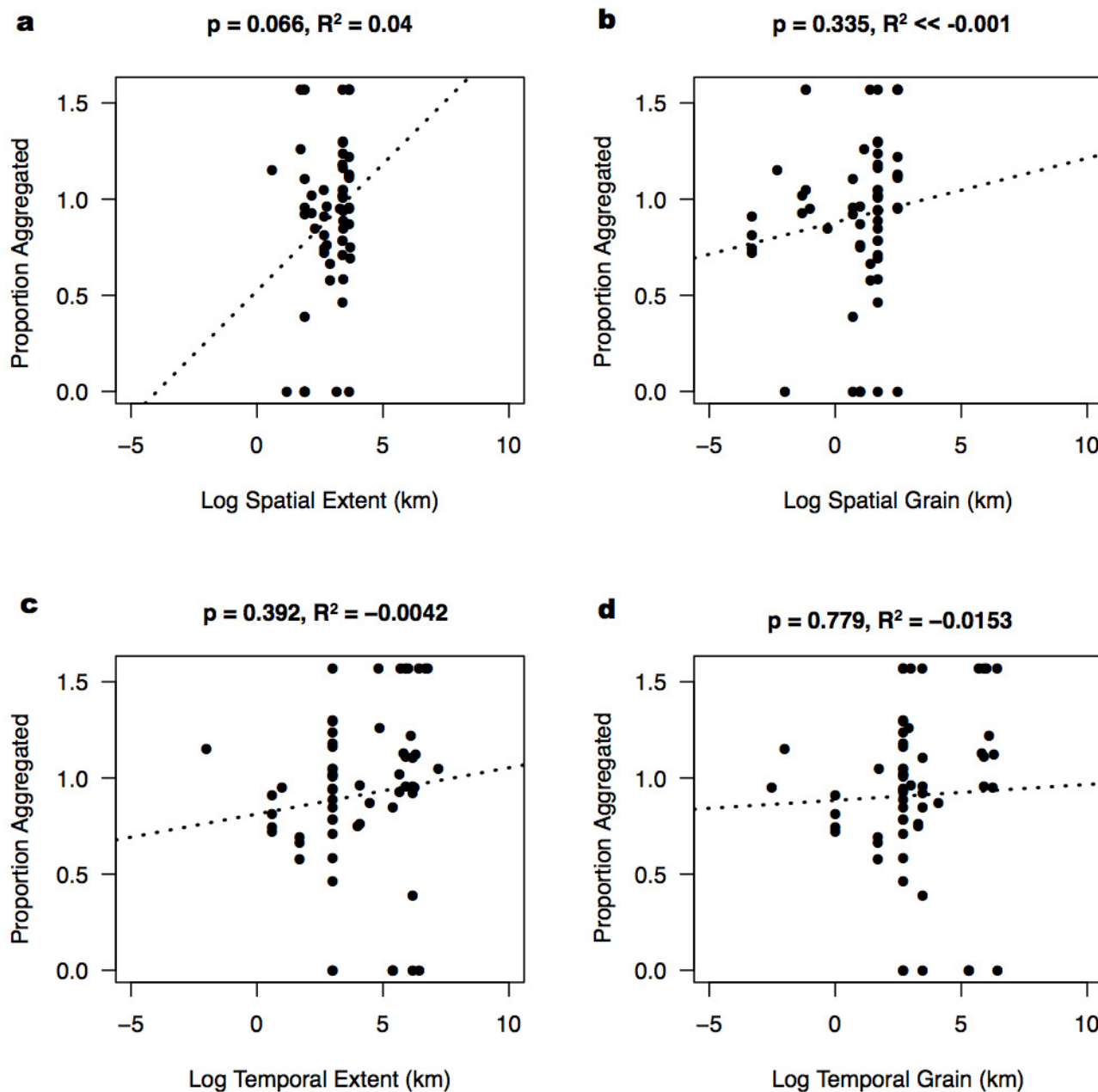
Australia (dark grey), South America (dark blue), Africa (orange). Point shapes indicate type of data: pollen (square), mammals (triangle), macroplants (circle). Data on terrestrial communities from ref. 2 are diamonds. Only mainland assemblages were included in the calculation for the weighted Loess curve and the density plots here and in Fig. 1.



**Extended Data Figure 4 | Results of PAIRS analyses of two Pleistocene–Holocene fossil data sets.** **a**, Mammal data for three periods: late Pleistocene (40,000–10,000 years ago), Holocene (10,000–500 years ago), and modern (present, literature survey data). Note the switch from >50% aggregated pairs to <50% aggregated pairs occurs in the Pleistocene versus Holocene data sets. **b**, Results for large and small mammals separately. There is a significant difference ( $P < 0.001$ ) between the Holocene and the Pleistocene for all mammals (blue bars) and for large mammals (purple bars) only ( $P = 0.015$ ). However, the direction of the shift was different. For all mammals, there were fewer positive associations in the Holocene, whereas, for large mammals only, there were more positive associations in the Holocene. **c**, North American pollen data from the past 21,000 years

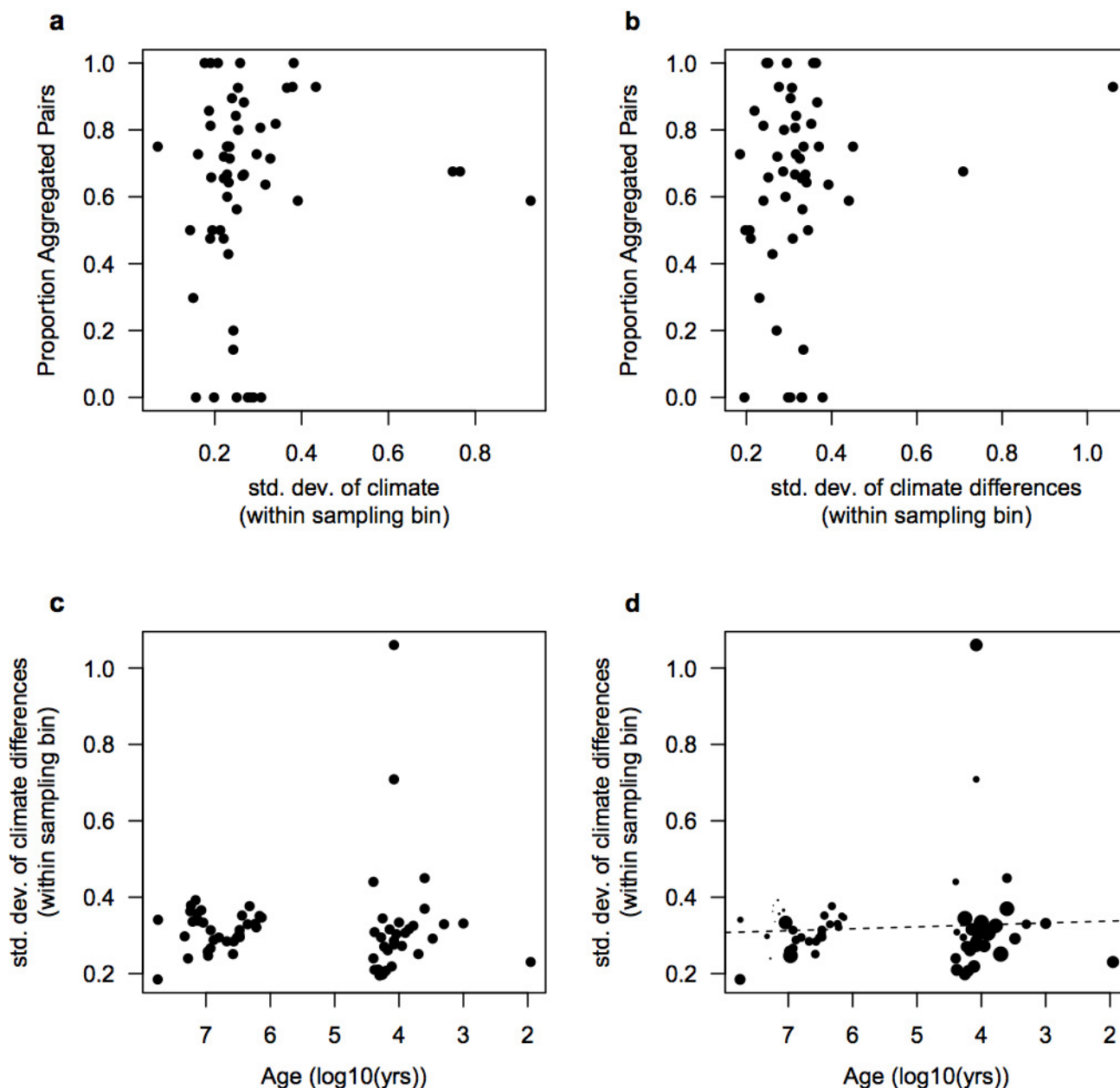
(modified from ref. 20). Data are from cores resolved into 1,000-year time slices. The size of the circle is related to the number of sites in the data set. The point at 0 represents a period from the present to 1,000 years ago, but is sampled from the top of the pollen cores using the same methodology as the older time slices. Note the trend of decreasing percentage of aggregations towards the present, especially in times with the largest numbers of sites (after 14,000 years). A weighted regression that takes into account the number of sites in each time slice is significant (dashed green line;  $P = 0.04$ , adjusted  $R^2 = 0.15$ ). The final time slice at 0 records a shift from a dominance of aggregated pairs to a dominance of segregated pairs. The sampling methods and data structure are the same for all time slices. Grey dashed line is at 50% in each panel.





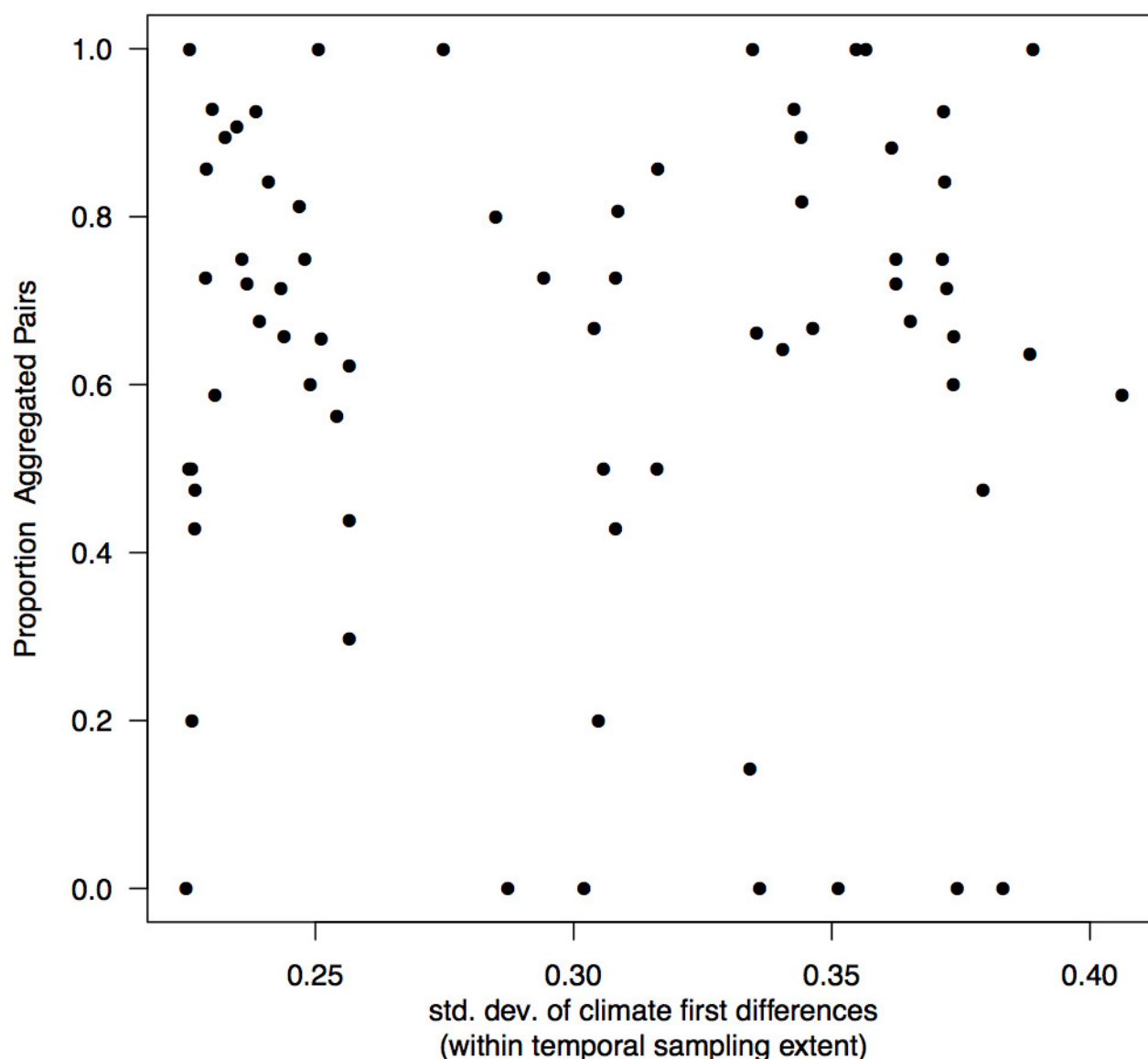
**Extended Data Figure 5 | Relationship between the proportion of aggregated pairs and scale.** The proportion of significant pairs that are aggregated is not the result of temporal or spatial scale of data. Arcsine transformation of the proportion of significant pairs that are aggregated

plotted as a function of spatial (a, b) or temporal (c, d) grain (b, d) or extent (a, c). Linear regressions are non-significant and adjusted  $R^2$  values are extremely low.



**Extended Data Figure 6 | Climate variability measured during the temporal extents of the fossil data sets.** Proportion of significant pairs that are aggregated shows no relationship with climate variability within a time interval. **a, b**, Climate variability was quantified as the standard deviation of all climate proxy data for that time interval (**a**), or the standard deviation of the first differences in climate across the interval (**b**). **c, d**, Climate variability (standard deviation of first differences) and

age of data sets show no relationship (**c**), suggesting no trend in climate variability sampled by the fossil data sets across the Phanerozoic. There is a significant, but weak, positive relationship (**d**, dashed line) between climate variability and decreasing age of the data sets when the linear model is weighted by sample size of climate proxy data enveloped by the temporal window of the fossil data sets ( $P = 0.007$ , adjusted  $R^2 = 0.0998$ ).



**Extended Data Figure 7 | Relationship between proportion of aggregated pairs and fixed-width time intervals.** High-amplitude Pleistocene climate variability oscillating between glacial and interglacial cycles may have imposed its own novel pressures on floral and faunal communities. Furthermore, ecological impacts may lag behind climate episodes themselves, complicating efforts to quantify climatic links to changes in the proportion of aggregated species pairs over time. Thus, limiting our measure of climate variability to the temporal span of the data

sets themselves may potentially not account for important (and possibly ecologically significant) climatic variability from the previous millennia. To incorporate this possibility, we re-analysed the relationship between the proportion of aggregated species pairs and climate variability of each data set, but included climate across the preceding 100,000 years, 10,000 years (not shown), and 1,000 years (not shown). As in the more restrictive analysis (Fig. 3a), there is consistently no relationship between climate variability and the proportion of aggregated species pairs.

Extended Data Table 1 | Raw data for Fig. 1

Dataset	Data Owner	Type	Cont	#Rand	#Agg	#Seg	#Spp	#Sites	Age (yr)	Temp Grain (yr)	Temp Extent (yr)	Spat Grain (km)	Spat Extent (km)
Turkana-Natoo	Behrensmeyer	M	AF	253	0	0	23	10	1365000	200000	250000	10	80
Turkana-Okote	Behrensmeyer	M	AF	1081	0	0	47	16	1470000	200000	250000	10	80
Turkana-Kaitio	Behrensmeyer	M	AF	351	0	0	27	8	1650000	200000	250000	10	80
Turkana-KBS	Behrensmeyer	M	AF	2015	0	1	64	37	1715000	200000	250000	10	80
Turkana-Kalochoro	Behrensmeyer	M	AF	276	0	0	24	7	2100000	200000	250000	10	80
Turkana-Burgi	Behrensmeyer	M	AF	1533	0	7	56	23	2250000	200000	250000	10	80
Turkana-Lomekwi	Behrensmeyer	M	AF	990	0	0	45	10	2985000	200000	250000	10	80
Turkana-TuluBor	Behrensmeyer	M	AF	778	0	2	40	12	3035000	200000	250000	10	80
Turkana-LokochoMorti	Behrensmeyer	M	AF	496	0	0	32	7	3705000	200000	250000	10	80
Sivallik-3.2-3.6Ma	Behrensmeyer	M	EA	210	0	0	21	16	3300000	3000	1500000	5	80
Sivallik-6.5Ma	Behrensmeyer	M	EA	820	0	0	41	30	6300000	3000	1500000	5	80
Sivallik-7Ma	Behrensmeyer	M	EA	4460	4	1	95	111	7600000	3000	1500000	5	80
Sivallik-8Ma	Behrensmeyer	M	EA	5050	0	0	101	170	8500000	3000	1500000	5	80
Sivallik-9Ma	Behrensmeyer	M	EA	4186	0	0	92	177	9400000	3000	1500000	5	80
Sivallik-11-10Ma	Behrensmeyer	M	EA	10703	4	24	147	92	11000000	3000	1500000	5	80
Sivallik-13-12Ma	Behrensmeyer	M	EA	8376	6	3	130	168	13100000	3000	1500000	5	80
Sivallik-15-14Ma	Behrensmeyer	M	EA	4454	7	4	95	38	14500000	3000	1500000	5	80
Sivallik-17-16Ma	Behrensmeyer	M	EA	2210	0	1	67	20	17100000	3000	1500000	5	80
Eurasia-spp-MN17	Eronen	M	EA	3644	9	2	86	32	2750000	650000	650000	300	4500
Eurasia-spp-MN16	Eronen	M	EA	15545	25	6	177	53	3000000	800000	800000	300	4500
Eurasia-spp-MN15	Eronen	M	EA	5993	2	0	110	27	3800000	800000	800000	300	4500
Eurasia-spp-MN14	Eronen	M	EA	2145	0	0	66	28	4750000	1100000	1100000	300	4500
Eurasia-spp-MN11	Eronen	M	EA	14532	2	1	171	44	8450000	800000	800000	300	4500
Eurasia-spp-MN10	Eronen	M	EA	8514	1	0	131	42	9250000	500000	500000	300	4500
Eurasia-spp-MN07	Eronen	M	EA	9977	30	4	142	65	11850000	1300000	1300000	300	4500
Eurasia-spp-MN06	Eronen	M	EA	9581	10	0	139	87	13850000	2700000	2700000	300	4500
Eurasia-spp-MN05	Eronen	M	EA	14119	51	26	169	96	16100000	1800000	1800000	300	4500
Eurasia-spp-MN04	Eronen	M	EA	7994	7	0	127	65	17500000	1000000	1000000	300	4500
Eurasia-spp-MN03	Eronen	M	EA	4803	39	9	99	57	19000000	2000000	2000000	300	4500
Eurasia-spp-MN02	Eronen	M	EA	3402	0	1	83	35	21400000	2800000	2800000	300	4500
SAfr-LgMammals-MIS1	Faith	M	AF	1091	25	12	48	103	12000	1000	12000	10	600
SAfr-LgMammals-MIS2	Faith	M	AF	906	19	21	44	28	24000	2000	12000	10	600
NAm-mammals-Modern	Lyons	M	NA	9267	541	777	146	67	30	50	50	50	5000
NAm-mammals-Holo	Lyons	M	NA	9860	404	467	147	214	5000	2000	10000	10	5000
NAm-mammals-Pleist	Lyons	M	NA	12727	667	467	167	88	25000	12500	30000	10	4500
crypticTaxaDesertRodents	Riddle	M	NA	1832	39	20	62	171	7.5	0.003	10	0.1	2000
African-SPP-MOD	Tóth	M	AF	15795	52	84	179	14	30	50	50	25	800
African-SPP-HIST	Tóth	M	AF	15632	36	85	178	14	90	50	50	25	800
WI-overstory-SPP-1950	Waller	PI	NA	512	26	23	34	168	60	1	4	0.0005	475
WI-understory-SPP-1950	Waller	PI	NA	12726	596	706	168	152	60	1	4	0.0005	475
WI-overstory-SPP-2000	Waller	PI	NA	688	33	20	39	168	10	1	4	0.0005	475
WI-understory-SPP-2000	Waller	PI	NA	12455	400	511	164	152	10	1	4	0.0005	475
Abo	DiMichele	PI	NA	934	9	3	44	207	290000000	55	16000000	0.07	450
PennWolf	DiMichele	PI	NA	1429	2	0	54	28	300000000	500	5000000	0.07	80
Rotliegend	DiMichele	PI	EA	988	2	0	45	62	300000000	1000	6000000	0.07	72
Calhoun Coal	DiMichele	PI	NA	378	0	0	28	55	304000000	500	10000	0.04	50
Kootenai	Jud	PI	NA	494	0	2	32	17	115000000	500	1000	0.01	15
Molteno5	Labandeira	PI	AF	3655	0	0	94	8	234000000	3000	250000	0.5	200
Molteno4	Labandeira	PI	AF	2278	0	0	68	12	234500000	3000	250000	0.5	200
Molteno3	Labandeira	PI	AF	8778	0	0	133	20	235000000	3000	250000	0.5	200
Molteno2	Labandeira	PI	AF	6312	9	7	113	43	235500000	3000	250000	0.5	200
Molteno1	Labandeira	PI	AF	8256	0	0	129	9	236000000	3000	250000	0.5	200
Burgersdorp	Labandeira	PI	AF	300	0	0	25	8	244000000	3000	250000	0.5	200
EarlyEocenePlants	Wing	PI	NA	3086	99	55	81	78	55600000	500	460000	0.05	150
LatePaleocenePlants	Wing	PI	NA	2101	32	12	66	42	56200000	500	460000	0.05	150
BigCedarRidgePlants	Wing	PI	NA	8238	15	3	129	100	73000000	0	0	0.005	4
pollenPG	Bercovici	Po	NA	3808	20	0	88	150	66027000	2865	68000	25	55
pollenK	Bercovici	Po	NA	4041	49	5	91	71	66065000	825	73000	15	55
NorthAm-pollen-0ka	Blois	Po	NA	3501	21	48	85	252	60	500	1000	50	2645.993
NorthAm-pollen-1ka	Blois	Po	NA	3522	27	21	85	445	1000	500	1000	50	2645.993
NorthAm-pollen-2ka	Blois	Po	NA	3541	19	10	85	438	2000	500	1000	50	2645.993
NorthAm-pollen-3ka	Blois	Po	NA	3286	21	14	82	410	3000	500	1000	50	2645.993
NorthAm-pollen-4ka	Blois	Po	NA	3285	27	9	82	397	4000	500	1000	50	2625.521
NorthAm-pollen-5ka	Blois	Po	NA	3202	25	13	81	397	5000	500	1000	50	2625.521
NorthAm-pollen-6ka	Blois	Po	NA	3472	10	4	84	354	6000	500	1000	50	2625.521
NorthAm-pollen-7ka	Blois	Po	NA	3141	16	3	80	335	7000	500	1000	50	2625.521
NorthAm-pollen-8ka	Blois	Po	NA	2899	25	2	77	300	8000	500	1000	50	2623.018
NorthAm-pollen-9ka	Blois	Po	NA	2901	18	7	77	280	9000	500	1000	50	2623.018
NorthAm-pollen-10ka	Blois	Po	NA	2528	21	7	72	251	10000	500	1000	50	2554.072
NorthAm-pollen-11ka	Blois	Po	NA	2466	17	2	71	205	11000	500	1000	50	2554.072
NorthAm-pollen-12ka	Blois	Po	NA	2542	13	1	72	155	12000	500	1000	50	2554.072
NorthAm-pollen-13ka	Blois	Po	NA	2542	12	2	72	117	13000	500	1000	50	2532.103
NorthAm-pollen-14ka	Blois	Po	NA	2200	8	3	67	76	14000	500	1000	50	2532.103
NorthAm-pollen-15ka	Blois	Po	NA	1946	3	4	63	50	15000	500	1000	50	2532.103
NorthAm-pollen-16ka	Blois	Po	NA	1826	2	2	61	39	16000	500	1000	50	2532.103
NorthAm-pollen-17ka	Blois	Po	NA	1480	1	4	55	30	17000	500	1000	50	2532.103
NorthAm-pollen-18ka	Blois	Po	NA	1324	1	1	52	21	18000	500	1000	50	2532.103
NorthAm-pollen-19ka	Blois	Po	NA	1484	1	0	55	16	19000	500	1000	50	2531.385
NorthAm-pollen-20ka	Blois	Po	NA	1481	0	4	55	15	20000	500	1000	50	1481.922
NorthAm-pollen-21ka	Blois	Po	NA	1176	0	0	49	13	21000	500	1000	50	2134.789

Numbers of aggregated versus segregated pairs and spatial and temporal scale of the ETE data sets included in this analysis. M, mammals; PI, macroplants; Po, pollen. AF, Africa; EA, Eurasia; NA, North America. #Rand, the number of taxon pairs that were not significantly different from random. #Agg, the number of significant taxon pairs that were aggregated. #Seg, the number of significant taxon pairs that were segregated. #Spp, the number of species in the data set. #Sites, the number of sites in the data set. Age (yr) is the midpoint age of the data set. Temp Grain (yr), temporal grain in years or the average amount of time encompassed by a site in the data set. Temp Extent (yr), the maximum amount of time encompassed by a data set. Spat Grain (km), the average distance from a site that fossils were transported. Spat Extent (km), the maximum linear distance encompassed by the data set.



Extended Data Table 2 | Effect of taxonomic resolution

Dataset	% Aggregations for Species	% Aggregations for Genera	Direction of difference when lumping
WI Overstory Plants-1950	0.531	0.800	Increase
WI Overstory Plants-2000	0.622	0.773	Increase
Eurasia mammals MN02	0	0	No change
Eurasia mammals MN03	0.815	0.732	Decrease
Eurasia mammals MN04	1	0.75	Decrease
Eurasia mammals MN05	0.662	0.608	Decrease
Eurasia mammals MN06	1	0.75	Decrease
Eurasia mammals MN07	0.882	0.864	Increase
Eurasia mammals MN09	1	0.75	Decrease
Eurasia mammals MN10	0.667	0.857	Increase
Eurasia mammals MN11	NA	1	Increase
Eurasia mammals MN14	NA	NA	No change
Eurasia mammals MN15	1	1	No change
Eurasia mammals MN16	0.806	1	Increase
Eurasia mammals MN17	0.818	0.619	Decrease
Africa mammals - historical	0.298	0.028	Decrease
Africa mammals – modern	0.382	0.031	Decrease
Great Basin Rodents Cryptic	0.661	0.500	Decrease

Change in the proportion of significant pairs when data sets are analysed at the species and genus levels. If lower taxonomic resolution of fossil data sets is driving the pattern of increased aggregations in the fossil data, we would expect to see increases in the percentage of aggregations when data are analysed at the genus level. Instead, most data sets show a decrease in the percentage of aggregated pairs. Only 6 of the 18 data sets analysed at multiple taxonomic resolutions show the expected increase. Nine show a decrease and three show no change. One data set (Great Basin Rodents Cryptic) was analysed at the species level and then taxonomically resolved with genetic data to include cryptic species. For that data set only, '% aggregations for genera' corresponds to the data set with cryptic species lumped and '% aggregations for species' corresponds to the data set with cryptic species resolved.

Extended Data Table 3 | Proportion of aggregated pairs across critical intervals

Taxon	Event	Place	Trend in % Positive Associations	p value
Large Mammals	Pleistocene-Holocene Transition	South Africa	Increase	0.007
Mammals	Pleistocene-Holocene Transition	North America	Decrease	<0.001
Pollen	Pleistocene-Holocene Transition	North America	Decrease	0.0008
Plants	PETM	North America	No Change	0.556
Pollen	K-Pg	North America	No Change	0.853
Pollen	Permian Crises	Greenland	No significant pairs	na

Significance of change in positive versus negative associations across critical intervals.

# Influence of extreme weather disasters on global crop production

Corey Lesk<sup>1</sup>, Pedram Rowhani<sup>2</sup> & Navin Ramankutty<sup>1,3</sup>

**In recent years, several extreme weather disasters have partially or completely damaged regional crop production<sup>1–5</sup>. While detailed regional accounts of the effects of extreme weather disasters exist, the global scale effects of droughts, floods and extreme temperature on crop production are yet to be quantified. Here we estimate for the first time, to our knowledge, national cereal production losses across the globe resulting from reported extreme weather disasters during 1964–2007. We show that droughts and extreme heat significantly reduced national cereal production by 9–10%, whereas our analysis could not identify an effect from floods and extreme cold in the national data. Analysing the underlying processes, we find that production losses due to droughts were associated with a reduction in both harvested area and yields, whereas extreme heat mainly decreased cereal yields. Furthermore, the results highlight ~7% greater production damage from more recent droughts and 8–11% more damage in developed countries than in developing ones. Our findings may help to guide agricultural priorities in international disaster risk reduction and adaptation efforts.**

In many regions of the world, there have been considerable changes in the nature of droughts, floods and extreme temperature events since the middle of the twentieth century<sup>6–8</sup>. Over agricultural areas, disasters arising from extreme weather can cause marked damage to crops and food system infrastructure, with the potential to destabilize food systems and threaten local to global food security. In recent years, nearly one-quarter of all damage and losses from climate-related disasters has been in the agricultural sector in developing countries<sup>9</sup>. With such disasters expected to become more common in the future<sup>1,6,7</sup>, policymakers need robust scientific information to develop effective disaster risk management and adaptation interventions (for example, infrastructure, technology, management and insurance) to protect the most vulnerable populations and to ensure global food security.

Whether an extreme weather event results in a disaster depends not only on the severity of the event itself, but also on the vulnerability and exposure of the human and natural systems that experience it<sup>6</sup>. Past research has addressed agricultural effects of specific weather extremes with fixed definitions, such as degree days above some threshold<sup>10–15</sup>. This approach probably underestimates the crop effects of extreme weather disasters (EWDs), because similar extreme weather events may have differing effects depending on the vulnerability of the exposed system.

In this study, we address this bias by using a disaster data set compiled based on human impact. In addition, we attend to two further limitations of previous work on extreme weather and agriculture. First, several regional empirical studies have highlighted the adverse effects of extreme heat events on crop yields<sup>10–13</sup>, and global modelling efforts have estimated future crop yield declines due to increasing extreme heat stress<sup>14,15</sup>. But this emphasis on crop yields offers an incomplete picture of crop production because of the potential for compensation or compounding of yield impacts by changes in harvested area<sup>16</sup>, and because crop production (and not yields)—together with access

and utilization—determines food security<sup>2,4,7,17,18</sup>. Second, we seek to investigate the agricultural effects of often-overlooked extreme weather events, namely floods and extreme cold<sup>2,3</sup>. Thus, our study is the first, to our knowledge, that takes an empirical approach to estimating the influence of EWDs on crop area, yields and production at the global scale.

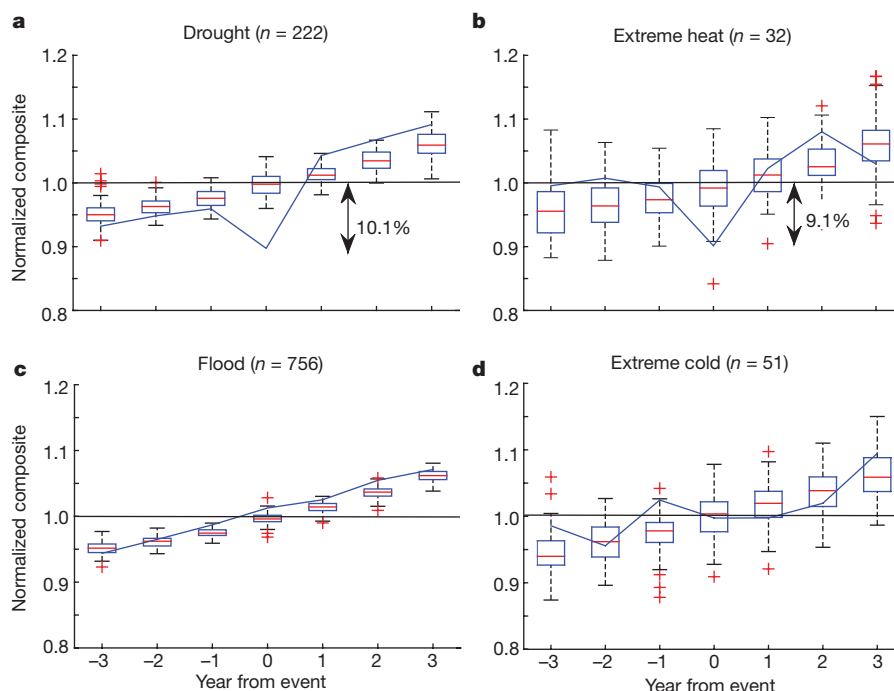
We use a statistical method, superposed epoch analysis<sup>19</sup> (also known as compositing, see Methods), to estimate average national per-disaster cereal production losses (Food and Agriculture Organization of the United Nations (FAO), <http://faostat3.fao.org>) across the globe due to reported droughts, floods and extreme temperature disasters from 1964 to 2007. Furthermore, we estimate the effects on cereal yield and harvested area separately to identify processes leading to production losses. On the basis of ~2,800 reported extreme hydro-meteorological disasters collated by the Emergency Events Database (EM-DAT, <http://www.emdat.be/database>), we find that national cereal production during a drought was significantly reduced by 10.1% on average (95% confidence interval 9.9–10.2%), while years with extreme heat led to national production deficits of 9.1% (8.4–9.5%; Fig. 1a, b and Extended Data Table 1). These production deficits were equivalent to roughly 6 years of production growth; however, no significant lasting effects were noted in the years after the disasters. Estimated mean production losses were driven mainly by a preponderance of disasters with moderate effects on crops, as opposed to a few extreme cases (Extended Data Fig. 1), and were not strongly influenced by sample size (see Extended Data Fig. 2, Extended Data Table 2 and Supplementary Discussion).

During 1964–2007, these estimated EWD effects represent a loss of 1,820 million Mg due to droughts (approximately equal to the global maize and wheat production in 2013), and 1,190 million Mg due to extreme heat disasters (more than the global 2013 maize harvest). Over 2000–2007 (the period with the most complete disaster reporting compared with earlier decades), 6.2% of total global cereal production was lost due to EWDs relative to an estimated counterfactual global production without EWD effects (3.0% to extreme heat and 3.2% to drought).

Cereal yield declines during EWDs were 5.1% (4.9–5.2%) and 7.6% (7.0–8.1%) for drought and extreme heat, respectively (Fig. 2a). The harvested area dropped 4.1% (4.0–4.3%) during droughts, but was not significantly affected by extreme heat (Fig. 2b). This may be due to the shorter duration of extreme heat relative to droughts—while approximately one-third of droughts in this study spanned several years, all extreme heat disasters took place within a single year. Droughts may thus be more likely to last long enough to cause complete crop failure and discourage planting, while extreme heat disasters, especially outside key crop developmental stages, may affect crop growth and reduce yields without critically damaging harvests.

Our estimated yield deficits from EWDs cannot be directly compared to previous studies of the impact of seasonal mean climate trends over the same period<sup>20</sup> (see Supplementary Discussion). However, we derived a comparable measure to that reported previously<sup>21</sup>, and estimated a yield sensitivity of 6–7% per 1 °C increase in seasonal mean

<sup>1</sup>Department of Geography, McGill University, Montreal H3A 0B9, Canada. <sup>2</sup>Department of Geography, University of Sussex, Brighton BN1 9QJ, UK. <sup>3</sup>Liu Institute for Global Issues and Institute for Resources, Environment and Sustainability, University of British Columbia, Vancouver V6T 1Z2, Canada.



**Figure 1 | Influence of EWDs on national cereal production.** **a–d**, Normalized production composites for drought ( $n = 222$ ) (**a**), extreme heat ( $n = 32$ ) (**b**), flood ( $n = 756$ ) (**c**) and extreme cold disasters ( $n = 51$ ) (**d**) over 7-year windows centred on the disaster year (blue lines). Box plots depict the distributions of 1,000 false-disaster control composites, with red crosses denoting extreme outliers, and red

dashes denoting medians. Production during drought and extreme heat years was 10.1% and 9.1% below the control mean, respectively, whereas no significant production signal was detected for floods or extreme cold. Production resumed normal levels immediately after drought and extreme heat. The increasing trend in production over the 7-year window reflects the observed growth trend.

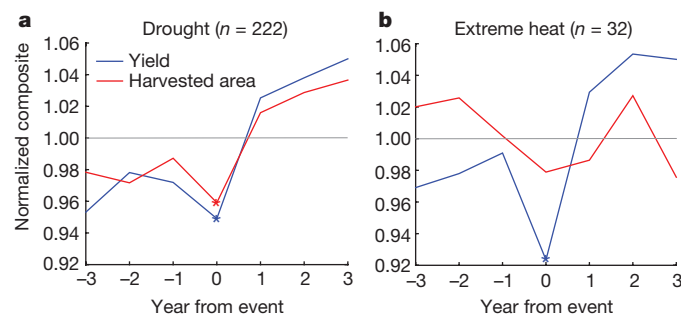
weather associated with extreme heat disasters, which suggests that our observed extreme heat effects are not necessarily independent from those detected in studies examining changes in seasonal temperatures (Extended Data Fig. 3). Methodological differences and uncertainties prevent us from drawing strong conclusions based on this comparison. Our drought impacts, however, seem to be independent of previous estimates that used seasonal weather anomalies (see Supplementary Discussion).

Our results do not show significant production effects from extreme cold and floods (Fig. 1c, d). A potential explanation for this is that floods tend to occur in the spring in temperate regions as a result of snowmelt, and cold weather susceptibility in most agricultural regions is highest outside the growing season, which may render a sizeable portion of the flood and extreme cold disasters analysed in this study agriculturally irrelevant. The estimated lack of response may also be

an artefact of the spatial dimension of these disasters. While drought and extreme temperature affect broad regions, floods are a function of both weather and topography, and can be highly localized within a country<sup>22</sup>. Since this study uses country-level agricultural statistics, one may speculate that a more noticeable flood effect on sub-national production is masked at the national scale.

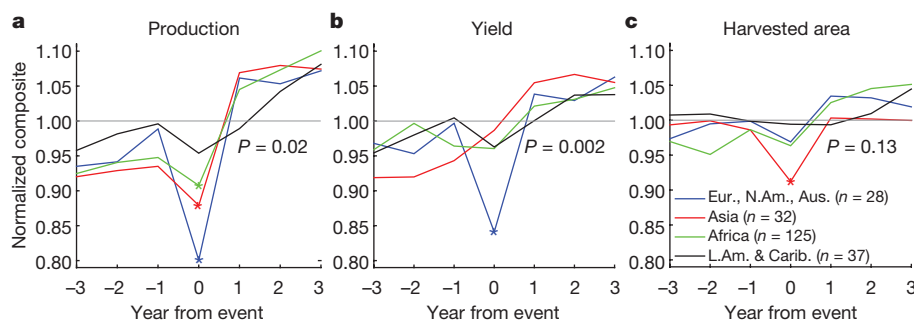
Several additional analyses offer more detailed insights into the effects of these EWDs on cereal production. Cereals in the more technically developed agricultural systems of North America, Europe and Australasia suffered most from droughts, facing on average a 19.9% production deficit compared to 12.1% in Asia, 9.2% in Africa, and no significant effect in Latin America and the Caribbean (overall difference in means  $P = 0.02$ ; Fig. 3a and Extended Data Tables 3 and 4). This more severe production impact in the developed nations was driven by a substantial yield deficit of 15.9%, with no significant reduction in harvested area (Fig. 3b, c). We see three possible explanations for this pattern. First, it may arise from a tendency among lower-income countries to encompass diverse crops and management across many small fields, which may allow for some fields to resist drought better than others. This might reduce the national drought sensitivity compared to higher-income countries, where large-scale monocultures are more dominant. Second, lower-income countries may better resist drought because smallholders tend to use risk-minimizing strategies compared to the yield-maximizing ones prevalent in higher-income countries. Third, the pattern may relate to generally lower fair-weather yields in lower-income countries. In Asia, we found a significant reduction of 8.8% in harvested area during droughts with no corresponding yield deficit, suggesting that this region has a greater tendency for total crop failure in the event of a drought rather than harvesting with reduced yields<sup>16</sup>. The production effects in Africa did not correspond to significant deficits in either yield or harvested area.

While the production of all three crops was similarly affected by droughts (5–6% deficit each; Fig. 4a, Extended Data Tables 2 and 5), only maize was significantly affected by extreme heat (11.7% deficit,  $P = 0.01$ ) (Fig. 4d). Maize was also the only crop with significant yield



**Figure 2 | Influence of EWDs on national cereal yields and harvested area.** **a, b**, Yield (blue) and harvested area (red) composites for drought ( $n = 222$ ) (**a**) and extreme heat ( $n = 32$ ) (**b**), with significant points (those lying beyond the control box plot whiskers) marked by asterisks (box plots not shown for clarity). Drought was associated with significant deficits in both yield and harvested area (5.1 and 4.1%), whereas extreme heat revealed only significant yield impacts of 7.6% with no significant effect on harvested area.





**Figure 3 | A regional analysis of the influence of drought.** a–c, Regional composites of production (a), yield (b) and harvested area (c) for drought, with significant points (those lying beyond the control box plot whiskers) marked by asterisks (box plots not shown for clarity). *P* values reflect significant differences between regions in drought-year response (Kruskal–Wallis test). The drought-year normalized production is 7.8%

effects (12.4%,  $P=0.002$ ) (Fig. 4b, e). We are hesitant to draw strong conclusions based on this difference, as it may be due to differing variance as well as mean (see Extended Data Table 6 and Supplementary Discussion). Furthermore, it may reflect the fact that maize is generally grown during summer months, which have the highest probabilities of extreme heat as defined in EM-DAT, while wheat is grown during the spring. Disaster data with monthly or daily resolution would enable us to investigate whether this apparent susceptibility of maize is a result of differing growing season.

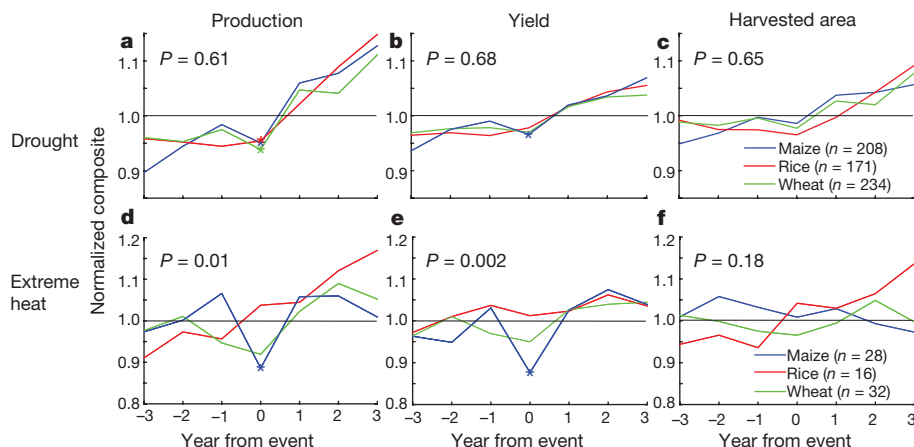
Finally, more recent droughts (1985–2007) caused cereal production losses averaging 13.7%, greater than the estimated 6.7% during earlier droughts (1964–1984) ( $P=0.008$ , Fig. 5), which may be due to any combination of rising drought severity (although whether drought severity has increased globally is presently debated)<sup>23–26</sup>, increasing vulnerability<sup>6,27</sup> and exposure to drought<sup>6</sup>, and/or changing reporting dynamics (Extended Data Fig. 4). Sample size limitations prevented us from repeating a regional and temporal analysis for extreme heat.

Some limitations of our analyses are worth noting. First, we mainly focus on four principal types of EWDs, but follow-up studies should include tropical storms and extreme precipitation and wind events, especially since they may have an increasingly important effect on agriculture in the context of climate change<sup>28</sup>. Second, our estimates are biased towards more recent disasters as they are more abundantly reported in EM-DAT than older ones (see Extended Data Fig. 4 and Supplementary Discussion). Third, we use EWDs from the EM-DAT

and 10.7% lower ( $P=0.02$ ) in developed Western countries ( $n=28$ ) than in Asia ( $n=32$ ) and Africa ( $n=125$ ) (a), a difference driven by a significantly greater yield deficit ( $P=0.002$ ) (b). Meanwhile, the Latin America (L.Am) and Caribbean (Carib.) region ( $n=37$ ) exhibits no significant response to drought. Aus., Australasia; Eur, Europe; N.Am., North America.

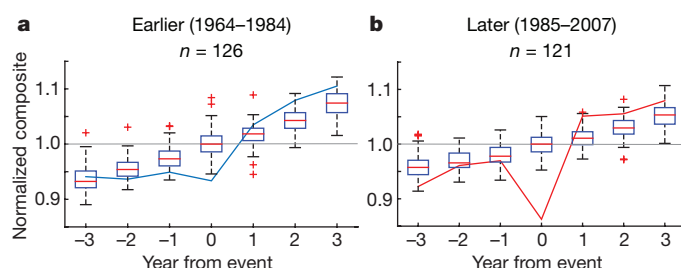
database, which collates disasters based on several criteria for substantial human impact (Methods). We may be underestimating the true effects of EWDs if disasters are included mainly based on urban impacts, or if extreme events occurring in sparsely populated areas are less likely to qualify as disasters. Finally, since we observe agricultural impacts at the national level, more notable local and regional effects of disasters may be muted (but conversely, finding a signal at the national level highlights the substantial influence of droughts and extreme heat). Future studies may arrive at a more detailed estimate by using subnational agricultural data, localizing the reported disasters within nations, selecting disasters taking place during the growing season, and controlling for severity of disasters. Linking the definitions of EWDs used in this study with statistical meteorological definitions will also enable a forecasting of future impacts.

Overall, there are four main conclusions from our study. First, over the period 1964–2007, drought and extreme heat substantially damaged national agricultural production across the globe. Within the framework of this study, no effect on agriculture was identified from floods and extreme cold. Second, drought reduced cereal yield and completely damaged crops, whereas extreme heat only affected yield, reflecting clear differences in the processes leading to overall production effects. Third, this study highlights an important temporal dimension to these impacts. While the damage to cereal production is considerable, this effect is only short term, as agricultural output rebounds and continues its growth trend after the disaster. Furthermore, we show that recent droughts had a larger effect on cereal production than earlier ones.



**Figure 4 | The influence of drought and extreme heat on maize, rice and wheat.** a–f, Drought and extreme heat composites of production (a, d), yield (b, e) and harvested area (c, f) for maize (blue), rice (red) and wheat (green), with significant points (those lying beyond the control box plot whiskers) marked by asterisks (box plots not shown for clarity). *P* values

reflect significant differences between crops in disaster-year response (Kruskal–Wallis test). Maize production ( $n=28$ ) responds more ( $P=0.01$ ) to extreme heat than wheat ( $n=32$ ) and rice ( $n=16$ ), an effect driven by a substantial yield deficit ( $P=0.002$ ). For drought data, maize ( $n=208$ ), rice ( $n=171$ ) and wheat ( $n=234$ ).



**Figure 5 | A temporal analysis of the influence of drought.**

**a, b,** Production composites for earlier (1964–1984,  $n = 126$ ) (a) and later (1985–2007,  $n = 121$ ) (b) droughts, with boxplots of 100 respective control composites. In later instances, mean drought-year production losses were greater (13.7%) than in earlier instances (6.7%;  $P = 0.008$ , Kruskal–Wallis test).

Finally, our regional and crop-specific analysis finds that developed nations suffer most from these EWDs.

Present climate projections suggest that extreme heat events will be increasingly common and severe in the future<sup>1</sup>. Droughts are likely to become more frequent in some regions, although considerable uncertainty persists in the projections<sup>6</sup>. This study, by highlighting the important historical effects of EWDs on agriculture, emphasizes the urgency with which the global cereal production system must adapt to extremes in a changing climate. Understanding the key processes leading to such crop losses enables an informed prioritization of disaster risk reduction and adaptation interventions to better protect the most vulnerable farming systems and the populations dependent on them.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 29 April; accepted 16 November 2015.**

- Battisti, D. S. & Naylor, R. L. Historical warnings of future food insecurity with unprecedented seasonal heat. *Science* **323**, 240–244 (2009).
- World Food Programme. *Pakistan – Flood Impact Assessment* <http://documents.wfp.org/stellent/groups/public/documents/ena/wfp225987.pdf> (2010).
- Gu, L. *et al.* The 2007 eastern US spring freeze: increased cold damage in a warming world. *Bioscience* **58**, 253 (2008).
- Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M. & García-Herrera, R. The hot summer of 2010: redrawing the temperature record map of Europe. *Science* **332**, 220–224 (2011).
- Coumou, D. & Rahmstorf, S. A decade of weather extremes. *Nature Clim. Change* **2**, 491–496 (2012).
- Intergovernmental Panel on Climate Change. in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (eds Field, C. B. *et al.*) 1–19 (Cambridge Univ. Press, 2012).
- United Nations Office for Disaster Risk Reduction. *The Pocket GAR 2013 – From Shared Risk to Shared Value: the Business Case for Disaster Risk Reduction* (UNISDR, 2013).
- World Meteorological Organization. *Atlas of Mortality and Economic Losses from Weather, Climate and Water Extremes (1970–2012)*. (WMO, 2014).

- Food and Agriculture Organization of the United Nations. *The Impact of Natural Hazards and Disasters on Agriculture and Food and Nutrition Security – A Call for Action to Build Resilient Livelihoods* (FAO, 2015).
- Lobell, D. B. *et al.* The critical role of extreme heat for maize production in the United States. *Nature Clim. Change* **3**, 497–501 (2013).
- Lobell, D. B., Sibley, A. & Ivan Ortiz-Monasterio, J. Extreme heat effects on wheat senescence in India. *Nature Clim. Change* **2**, 186–189 (2012).
- Lobell, D. B., Bänziger, M., Magorokosho, C. & Vivek, B. Nonlinear heat effects on African maize as evidenced by historical yield trials. *Nature Clim. Change* **1**, 42–45 (2011).
- Moriondo, M., Giannakopoulos, C. & Bindi, M. Climate change impact assessment: the role of climate extremes in crop yield simulation. *Clim. Change* **104**, 679–701 (2010).
- Teixeira, E. I., Fischer, G., van Velthuizen, H., Walter, C. & Ewert, F. Global hot-spots of heat stress on agricultural crops due to climate change. *Agric. For. Meteorol.* **170**, 206–215 (2013).
- Deryng, D., Conway, D., Ramankutty, N., Price, J. & Warren, R. Global crop yield response to extreme heat stress under multiple climate change futures. *Environ. Res. Lett.* **9**, 034011 (2014).
- Iizumi, T. & Ramankutty, N. How do weather and climate influence cropping area and intensity? *Glob. Food Sec.* **4**, 46–50 (2015).
- Johnstone, S. & Mazo, J. Global warming and the Arab Spring. *Survival* **53**, 11–17 (2011).
- Welton, G. *The Impact of Russia's 2010 Grain Export Ban* (Oxfam International, 2011).
- Brad Adams, J., Mann, M. E. & Ammann, C. M. Proxy evidence for an El Niño-like response to volcanic forcing. *Nature* **426**, 274–278 (2003).
- Lobell, D. B., Schlenker, W. & Costa-Roberts, J. Climate trends and global crop production since 1980. *Science* **333**, 616–620 (2011).
- Lobell, D. B. & Field, C. B. Global scale climate–crop yield relationships and the impacts of recent warming. *Environ. Res. Lett.* **2**, 014002 (2007).
- Thornton, P. K., Ericksen, P. J., Herrero, M. & Challinor, A. J. Climate variability and vulnerability to climate change: a review. *Glob. Chang. Biol.* **20**, 3313–3328 (2014).
- Dai, A. Increasing drought under global warming in observations and models. *Nature Clim. Change* **3**, 52–58 (2012).
- Sheffield, J., Wood, E. F. & Roderick, M. L. Little change in global drought over the past 60 years. *Nature* **491**, 435–438 (2012).
- Trenberth, K. E. *et al.* Global warming and changes in drought. *Nature Clim. Change* **4**, 17–22 (2014).
- Greve, P. *et al.* Global assessment of trends in wetting and drying over land. *Nature Geosci.* **7**, 716–721 (2014).
- Lobell, D. B. *et al.* Greater sensitivity to drought accompanies maize yield increase in the U.S. Midwest. *Science* **344**, 516–519 (2014).
- Gornall, J. *et al.* Implications of climate change for agricultural productivity in the early twenty-first century. *Phil. Trans. R. Soc. Lond. B* **365**, 2973–2989 (2010).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank R. Below, who is in charge of the EM-DAT project at the Centre for Research on the Epidemiology of Disasters, for sharing the data. We thank C. Champalle for testing the original idea using data over East Africa in a class project. This research was supported by a Discovery Grant from the Natural Science and Engineering Research Council of Canada to N.R.

**Author Contributions** This research was designed and coordinated by N.R. All authors performed analyses, discussed the results, and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.R. ([navin.ramankutty@ubc.ca](mailto:navin.ramankutty@ubc.ca)).

## METHODS

Superposed epoch analysis (SEA) is used to isolate an average EWD response signal using time series of national agricultural production data and EWDs. SEA is a statistical approach that has been used to enhance the signal (that is, influence of particular events) in time-series data, while reducing noise due to extraneous variables<sup>19</sup>. The EWDs are compiled from the Emergency Events Database (EM-DAT; <http://www.emdat.be/database>) and consist of 2,184 floods, 497 droughts, 138 extreme heat and 194 extreme cold disasters from 177 countries over the period 1964–2007. EM-DAT collects information on a reported disaster if at least ten people died, a state of emergency was declared, international assistance was called, or at least 100 people were injured, made homeless or required immediate assistance. Disaster reports are gathered from various organizations including United Nations agencies, governments, and the International Federation of Red Cross and Red Crescent Societies<sup>8</sup>. The agricultural data consist of country-level total production, average yield, and total harvested area data for 16 cereals (<http://faostat3.fao.org>) covering the 177 countries in the set of EWDs from 1961 to 2010.

From the time series of agricultural data, we extracted shorter sets of time series using a 7-year window centred on the year of occurrence of each EWD, with 3 years of data preceding and following each EWD. The data were normalized to the average of the 3 years preceding and following the event to remove the absolute magnitude of national data from the signal. For multi-year droughts, we averaged across all drought years to produce a single disaster year datum. For a 3-year drought, for example, the 7-year window became a 9-year window with seven data points (with the middle 3 years being averaged and assigned to year 0). The 7-year sets of EWD time series were then centred on the disaster year and averaged year-wise to yield single composited time-series of production, yield and harvested area for each EWD type (a total of 12 composited time series). The averaging thus strengthens the signal at the central year of EWD occurrence, while also cancelling the noise in the non-disaster years preceding and following the event.

During compositing, points on individual time series co-occurring with another disaster in the set were excluded from the mean. This procedure resulted in variable sample size across the 7 years of the composites. For brevity, we have presented mean sample sizes across all years; complete tabulated sample sizes are displayed in Extended Data Tables 2 and 4. Our composited mean estimate does not seem to be influenced by outliers (see Extended Data Fig. 1 and Supplementary Discussion). The signal-to-noise strength will certainly depend on the sample size, and we performed an analysis to estimate the influence of sample size (see Extended Data Tables 2 and 4, Extended Data Fig. 2 and Supplementary Discussion).

In addition to average per-disaster estimates, we also calculated aggregate production losses over specific time periods. For each extreme heat or drought, we first applied the average per-disaster percentage loss estimate (different values for extreme heat or drought) to the average national production across the six adjacent non-disaster years. We then computed the aggregate drought or heat-related global

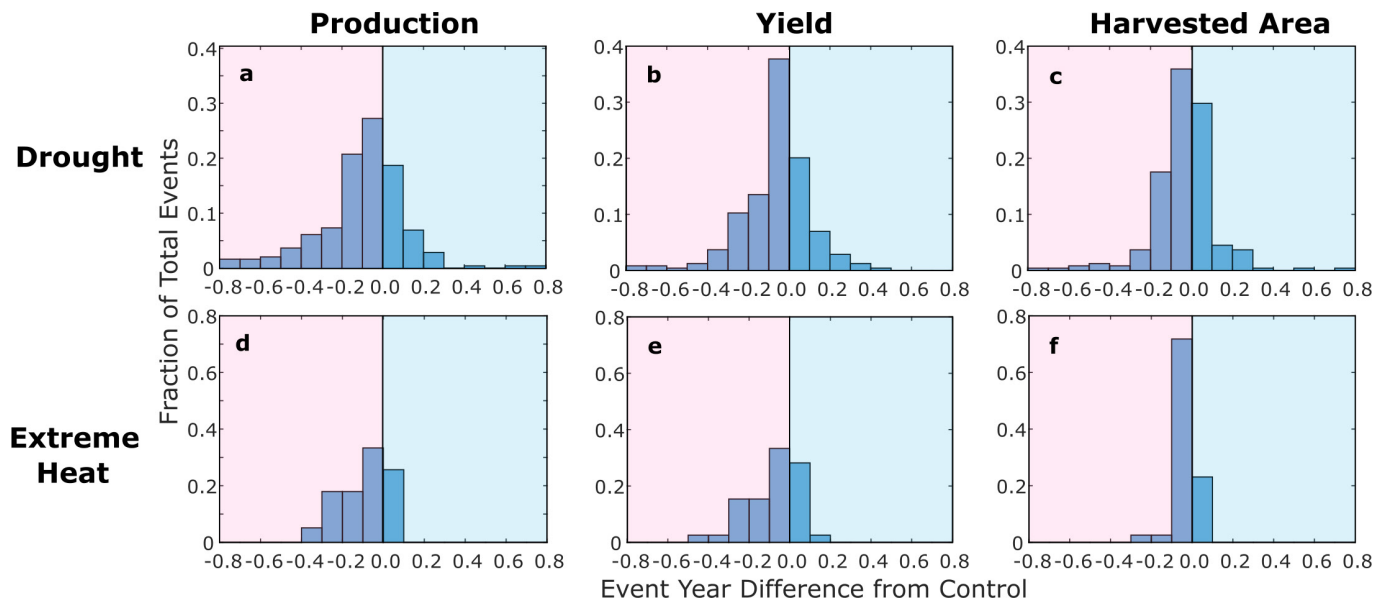
production loss for each year by summing the production losses for each disaster over the given time period. We estimated the percentage of global production lost to the EWDs relative to an estimated counterfactual global production in a world without EWDs (the latter being the sum of observed global production plus the estimated production loss).

The significance-testing procedure involved setting up a 'control' estimate by randomly resampling the agricultural data using sets of fictitious disasters with randomly generated years and countries of occurrence. The fictitious EWD time series were averaged as for the true ones to yield composited 'control' time series, and the entire process was repeated 1,000 times. We quantified EWD-year deficits in production, yield and harvested area by subtracting the true EWD time series from the mean of the controls. Excluding randomly generated disasters that happened to be real disasters systematically raised the impact estimates by ~1%; to present a more conservative and rigorous detection of the disaster signal, we elected not to exclude such pseudo-disasters. Note that we chose not to de-trend the time series before compositing to remove technology-driven growth, but rather simply estimate the disaster signal as difference from control (see Fig. 1). We estimated the 95% confidence intervals for our point estimates of impacts using an approach similar to a delete-one jackknife resampling method (see Supplementary Discussion).

The percentage significance of each estimate of the EWD composites relative to control was estimated as the percentage of 1,000 control points less than the EWD composite estimate for each year. Points with estimated significance of <0.5% or >99.5% were considered significant deficits and surpluses, respectively, corresponding to a two-tailed 99% confidence level. While we chose a two-tailed approach for robustness, we found no significant surpluses. The significant points appear as asterisks in Figs 2–4, while for Figs 1 and 5 we present the EWD composites with the distribution of controls represented as an array of box-and-whisker plots for a visual representation of significance. The complete tabulated percentage significance values are presented in Extended Data Tables 1, 3 and 5.

The earlier-versus-later analysis for droughts was performed by applying the SEA procedure to the set of droughts divided roughly equally into earlier and later halves. Similarly, the regional analysis was conducted by repeating SEA for full set of disasters divided into four regional groupings, and the by-crop composites were obtained by repeating SEA on the full disaster sets using crop-specific agricultural data from the FAO (<http://faostat3.fao.org>). Statistical significance of differences between crop-specific, regional and earlier-versus-later composites was assessed using the Kruskal–Wallis test. We applied a quadratic transformation to the data for comparison to equalize variance between groups (verified using Levene's test), and used non-parametric tests when comparing groups as normal assumptions were not met (see Supplementary Discussion).

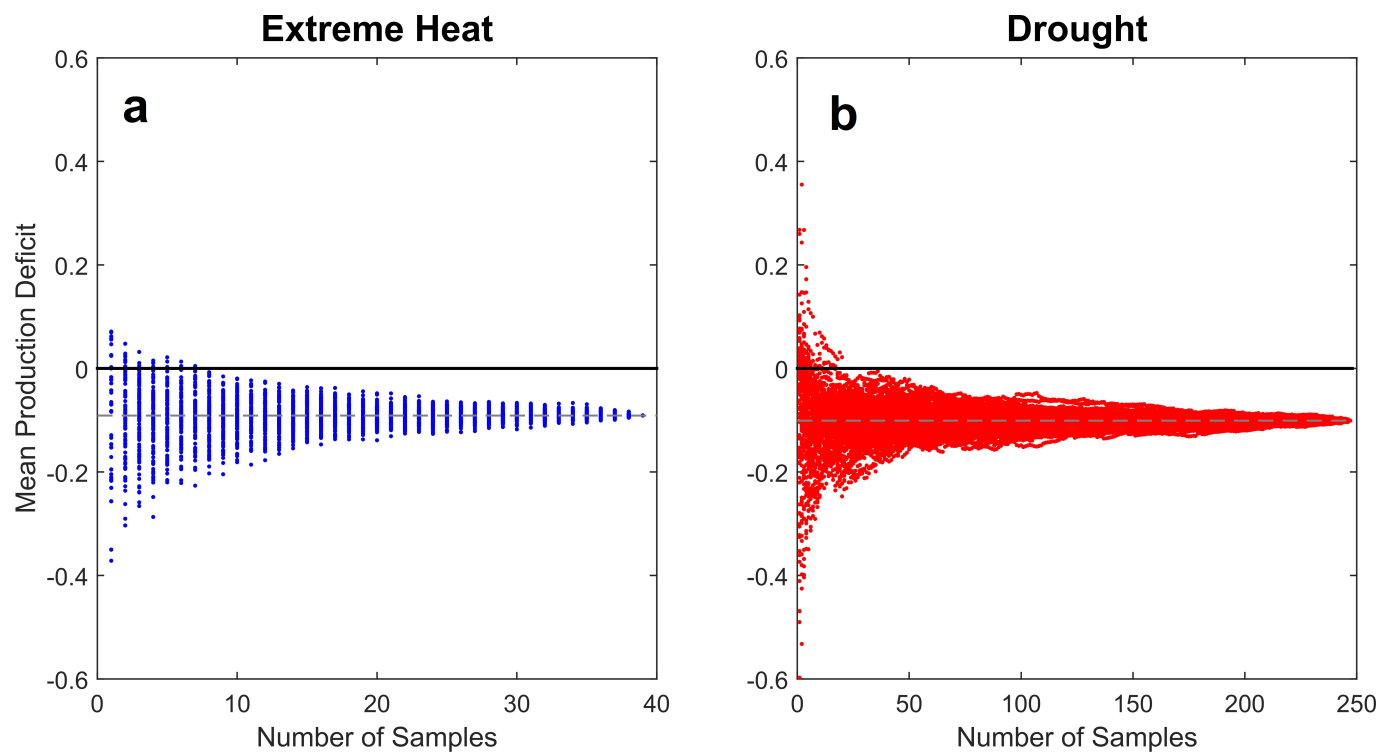
**Code availability.** All the core programs including codes to perform superposed epoch analysis and the various statistics described in this paper are available on Github (<https://github.com/nramankutty/SEA-code>).



**Extended Data Figure 1 | Distributions of individual responses to drought and extreme heat.** a–f, Histograms of disaster-year differences from means of 1,000 resampled controls for drought ( $n=222$ ) (a–c) and extreme heat ( $n=32$ ) (d–f). A preponderance of moderately negative

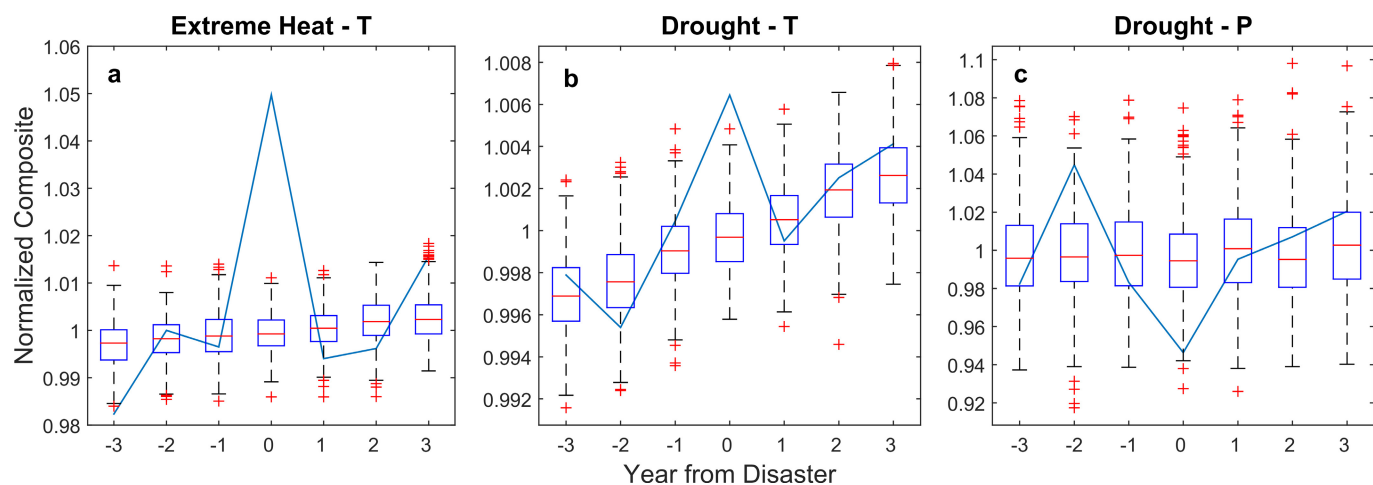
values (falling towards the right of the red shaded areas) underlies the negative mean disaster year signals, with a limited influence of extreme cases (those at the left of the red shaded areas).





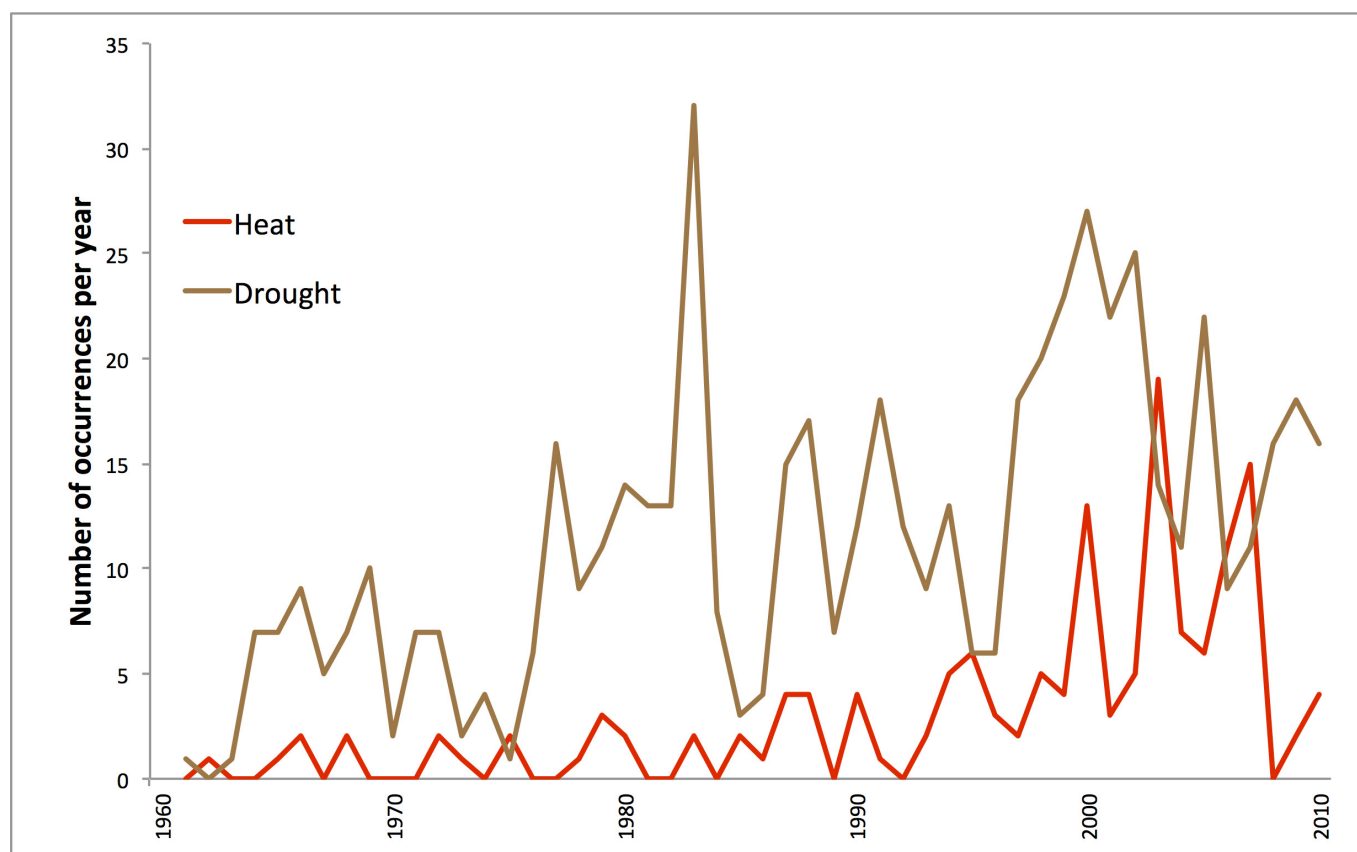
**Extended Data Figure 2 | The influence of sample size on estimated disaster effects. a, b,** Estimated mean 16-cereal aggregated production deficit for extreme heat (**a**) and drought (**b**) in 200 sub-samples with size of (1, 2, ...,  $n$ ) (points). Dotted grey line shows the final estimated mean

production deficit (9.1% for extreme heat, 10.1% for drought). Most of the initial variability at low sample sizes dissipates into the mean at well below the actual sample size ( $n = 39$  for extreme heat,  $n = 247$  for drought).



**Extended Data Figure 3 | Seasonal weather anomalies of drought and extreme heat disasters in EM-DAT.** a–c, Normalized composite mean growing season temperature for extreme heat ( $n=32$ ) (a) and drought ( $n=222$ ) (b), and total precipitation for drought (c). Box plots depict the distributions of 1,000 false-disaster control composites, with red crosses

denoting extreme outliers and red dashes denoting medians. Years with extreme heat correspond to seasonal temperature anomalies of  $1.2^{\circ}\text{C}$ , while drought years have only  $0.15^{\circ}\text{C}$  warmer temperatures, with no significant precipitation anomaly.



**Extended Data Figure 4 | Time series of the number of extreme heat and drought disasters per year from the EM-DAT database.** The EM-DAT database is based on a compilation of disaster reports gathered from various organizations including United Nations agencies, governments and the International Federation of Red Cross and Red Crescent Societies. The time

series of reported disasters per year exhibits an increasing trend, probably the result of more complete disaster reporting in more recent decades with a possible contribution from increasing disaster incidence. There is also large inter-annual variability in the number of disasters.

Extended Data Table 1 | Statistical significance of 16-cereal aggregate analysis

		Percent Significance (% of 1000 controls < disaster)						
Year from Disaster		-3	-2	-1	0	1	2	3
Drought	Production	13.5	16.8	9.7	0	98.9	98.4	92.9
	Yield	22.0	71.0	8.9	0.0	93.3	86.3	68.7
	Harvested Area	21.5	7.3	35.2	0.0	92.4	95.3	87.5
Extreme Heat	Production	86.2	87.2	70.5	0.0	61.3	85.8	27.2
	Yield	62.1	56.5	59.8	0.1	81.9	82.7	60.5
	Harvested Area	84.2	91.2	68.9	22.1	25.1	70.6	9.8
Extreme Cold	Production	84.5	44.0	97.2	48.7	31.5	30.5	81.9
	Yield	41.4	18.9	97.2	50.2	60.7	57.8	94.8
	Harvested Area	90.1	69.3	81.0	53.2	19.1	17.5	47.9
Flood	Production	37.1	67.7	90.6	98.6	93.2	97.2	73.6
	Yield	57.1	65.0	29.5	93.1	85.1	98.3	96.2
	Harvested Area	26.6	61.6	97.3	95.7	85.9	75.0	30.6

Percentage of points on control composites less than EWD composites for 16-cereal aggregate, 1,000 control replicates total.



Extended Data Table 2 | Sample sizes for individual crop and 16-cereal aggregate analyses

		n =							Mean
Year from Disaster		-3	-2	-1	0	1	2	3	
Drought	Wheat	139	129	144	146	132	122	123	134
	Rice	175	175	186	188	170	159	147	171
	Maize	205	207	225	231	206	196	183	208
Extreme Heat	Wheat	34	30	28	39	39	26	27	32
	Rice	18	14	15	19	17	16	14	16
	Maize	30	27	24	35	35	24	22	28
Drought	16-cereal aggregate	221	220	243	247	219	210	196	222
Extreme Heat	16-cereal aggregate	34	30	28	39	39	26	27	32
Extreme Cold	16-cereal aggregate	53	47	59	58	44	44	52	51
Flood	16-cereal aggregate	737	712	671	1345	625	602	603	756

Extended Data Table 3 | Statistical significance of regional analysis

		Percent Significance (% of 1000 controls < disaster)						
	Year from Disaster	-3	-2	-1	0	1	2	3
Production	Africa	23.1	24.1	11.4	0.0	93.8	96.0	92.1
	Asia	31.4	23.3	14.5	0.0	93.4	83.7	49.9
	North America, Europe, Australasia	27.8	18.9	61.4	0.0	96.3	80.9	82.2
	Latin America and Caribbean	48.2	65.0	74.0	7.6	24.2	65.4	75.4
Yield	Africa	16.4	85.7	3.6	0.7	88.0	86.2	88.4
	Asia	10.6	3.0	5.0	32.2	96.1	90.6	51.3
	North America, Europe, Australasia	73.9	18.2	71.8	0.0	85.7	40.8	68.0
	Latin America and Caribbean	42.4	71.2	90.2	2.1	24.3	75.7	38.6
Harvested Area	Africa	43.2	3.6	48.5	0.7	85.1	90.1	67.7
	Asia	64.2	70.3	49.0	0.0	60.6	49.5	38.8
	North America, Europe, Australasia	5.7	23.3	37.7	4.6	96.7	93.9	81.3
	Latin America and Caribbean	57.0	65.1	52.4	38.5	32.7	49.7	76.8

Percentage of points on control composites less than EWD composites for 16-cereal aggregate by region, 1,000 control replicates total.

Extended Data Table 4 | Sample sizes for regional analysis

Year from Disaster	n =							Mean
	-3	-2	-1	0	1	2	3	
North America, Europe, Australasia	28	27	25	34	27	26	30	28
Asia	39	31	35	36	28	31	23	32
Africa	117	120	144	139	129	117	110	125
Latin America and Caribbean	37	42	39	38	35	36	33	37

Extended Data Table 5 | Statistical significance of individual crop analysis

		Percent Significance (% of 1000 controls < disaster)							
		Year from Disaster	-3	-2	-1	0	1	2	3
Drought	Maize	Production	3.1	44.0	82.7	0.4	97.3	77.8	70.3
		Yield	7.5	73.1	76.9	0.1	59.6	53.9	72.1
		Harvested Area	7.1	24.3	84.3	19.5	97.4	80.0	54.6
	Rice	Production	53.1	27.5	1.1	0.1	45.1	94.8	96.4
		Yield	38.0	25.9	1.6	3.0	67.0	92.0	78.1
		Harvested Area	67.7	34.7	16.1	1.0	24.4	87.7	96.6
	Wheat	Production	64.9	34.6	49.0	0.2	94.4	54.2	89.4
		Yield	77.4	70.1	47.2	2.0	63.5	64.6	28.6
		Harvested Area	35.5	23.2	49.2	5.0	91.0	59.3	96.1
Extreme Heat	Maize	Production	70.1	81.2	95.8	0.2	72.7	44.3	6.8
		Yield	59.1	28.7	90.0	0.0	62.5	85.4	35.6
		Harvested Area	67.4	93.7	81.8	52.8	64.7	18.0	7.5
	Rice	Production	29.1	62.5	39.3	77.4	68.5	86.2	80.0
		Yield	51.1	78.2	90.1	62.2	60.5	82.6	38.5
		Harvested Area	26.1	35.8	14.9	80.4	66.9	78.6	86.4
	Wheat	Production	63.6	83.0	25.9	2.3	57.9	83.1	35.9
		Yield	60.5	90.8	34.1	4.4	75.6	63.7	49.3
		Harvested Area	64.6	51.9	28.8	10.3	35.4	79.4	25.0

Percentage of points on control composites less than EWD composites for individual crop analysis, 1,000 control replicates total.



Extended Data Table 6 | Kruskal–Wallis assumptions test results for group comparison analyses

Analysis	Figure	Data	Normal? (Anderson-Darling test)	Equal variance? (Levene's Absolute test)
Regional	3a	Production	No	Yes
	3b	Yield	No	Yes
	3c	Harvested Area	No	Yes
Individual Crop: Drought	4a	Production	No	Yes
	4c	Yield	No	Yes
	4e	Harvested Area	No	Yes
Individual Crop: Extreme Heat	4b	Production	No	Yes
	4d	Yield	No	No
	4f	Harvested Area	No	No
Earlier-Later Droughts	5	Production	No	Yes

# The calcium sensor synaptotagmin 7 is required for synaptic facilitation

Skyler L. Jackman<sup>1</sup>, Josef Turecek<sup>1</sup>, Justine E. Belinsky<sup>1</sup> & Wade G. Regehr<sup>1</sup>

It has been known for more than 70 years that synaptic strength is dynamically regulated in a use-dependent manner<sup>1</sup>. At synapses with a low initial release probability, closely spaced presynaptic action potentials can result in facilitation, a short-term form of enhancement in which each subsequent action potential evokes greater neurotransmitter release<sup>2</sup>. Facilitation can enhance neurotransmitter release considerably and can profoundly influence information transfer across synapses<sup>3</sup>, but the underlying mechanism remains a mystery. One proposed mechanism is that a specialized calcium sensor for facilitation transiently increases the probability of release<sup>2,4</sup>, and this sensor is distinct from the fast sensors that mediate rapid neurotransmitter release. Yet such a sensor has never been identified, and its very existence has been disputed<sup>5,6</sup>. Here we show that synaptotagmin 7 (Syt7) is a calcium sensor that is required for facilitation at several central synapses. In Syt7-knockout mice, facilitation is eliminated even though the initial probability of release and the presynaptic residual calcium signals are unaltered. Expression of wild-type Syt7 in presynaptic neurons restored facilitation, whereas expression of a mutated Syt7 with a calcium-insensitive C2A domain did not. By revealing the role of Syt7 in synaptic facilitation, these results resolve a longstanding debate about a widespread form of short-term plasticity, and will enable future studies that may lead to a deeper understanding of the functional importance of facilitation.

Several mechanisms for facilitation have been proposed (Extended Data Fig. 1). In the 'buffer saturation' model, high concentrations of presynaptic  $\text{Ca}^{2+}$  buffer capture incoming  $\text{Ca}^{2+}$  before it binds to the rapid synaptotagmin isoforms (1, 2 and 9) that trigger vesicle fusion at most synapses<sup>7</sup>. If the  $\text{Ca}^{2+}$  buffer saturates during the first action potential, more  $\text{Ca}^{2+}$  reaches release sites during subsequent action potentials, producing facilitation<sup>6,8</sup>. Yet many facilitating synapses lack sufficient presynaptic  $\text{Ca}^{2+}$  buffer to account for this form of facilitation<sup>9</sup>. Another theory suggests that a specialized  $\text{Ca}^{2+}$  sensor responds to the smaller, longer-lasting  $\text{Ca}^{2+}$  signals between action potentials<sup>4</sup>. In one scenario, this sensor modulates  $\text{Ca}^{2+}$  channels to produce use-dependent increases in  $\text{Ca}^{2+}$  influx<sup>10</sup>. Several candidate proteins have been proposed to act in this manner<sup>11,12</sup>, but increased  $\text{Ca}^{2+}$  influx cannot account for facilitation at most synapses<sup>13</sup>. Alternatively, an unidentified  $\text{Ca}^{2+}$  sensor could mediate facilitation by directly increasing the probability of release ( $p$ ).

Syt7 is located presynaptically, and binds  $\text{Ca}^{2+}$  with high affinity and slow kinetics<sup>14–16</sup>, making it a promising candidate sensor for the modest increases in residual  $\text{Ca}^{2+}$  that mediate facilitation. Previous studies suggest that Syt7 contributes to a slow phase of transmission known as asynchronous release<sup>17,18</sup>, and to  $\text{Ca}^{2+}$ -dependent recovery from depression<sup>19</sup>, but the role of Syt7 in facilitation was not examined because these studies used synapses with prominent depression that obscures facilitation. We therefore examined synaptic transmission at four facilitating synapses: Schaffer collateral synapses between hippocampal CA3 and CA1 pyramidal cells<sup>9</sup> (Fig. 1a), thalamocortical synapses between layer 6 cortical pyramidal cells and thalamic relay

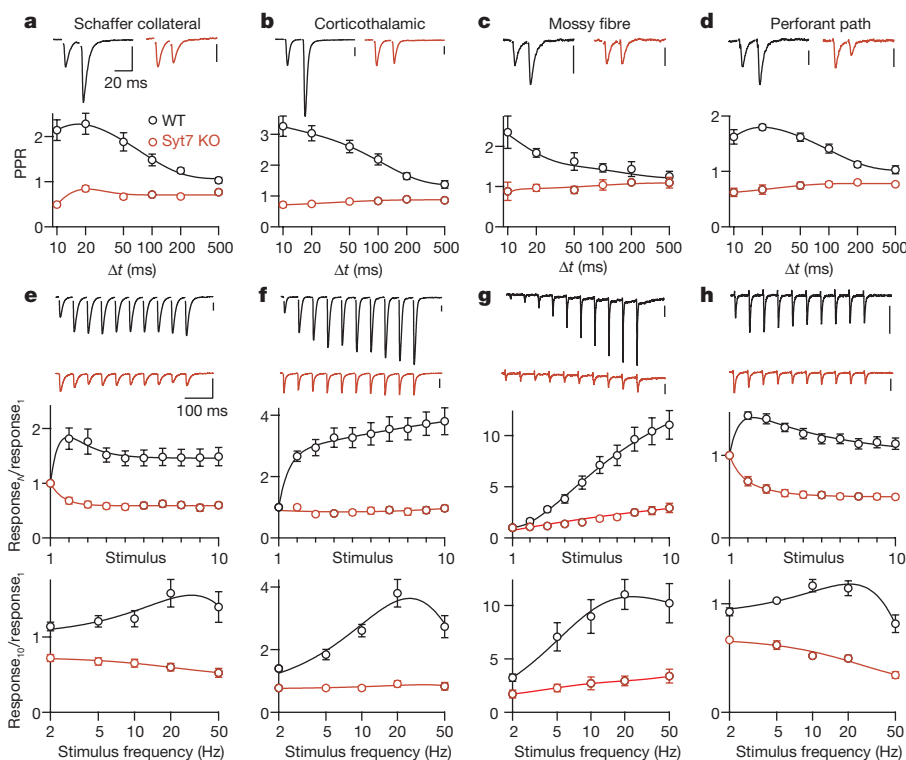
cells<sup>20</sup> (Fig. 1b), mossy fibre synapses between dentate granule and CA3 cells<sup>9</sup> (Fig. 1c), and perforant path synapses between layer II and III cells of the entorhinal cortex and dentate granule cells<sup>21</sup> (Fig. 1d). Immunohistochemistry shows that Syt7 is present in regions where these synapses are located (Extended Data Figs 2 and 3). Facilitation is often assessed using pairs of closely spaced stimuli. In slices from wild-type mice, paired-pulse facilitation resulted in ~2-fold enhancement of neurotransmitter release lasting several hundred milliseconds (Fig. 1a–d, black traces). In Syt7-knockout mice, paired-pulse facilitation was eliminated (Fig. 1a–d, red traces). Sustained high frequency activation produces up to tenfold enhancement in wild-type animals, but facilitation is eliminated in knockouts at all synapses except for mossy fibre synapses, where the remaining enhancement is consistent with use-dependent spike broadening that occurs at this synapse<sup>22</sup> (Fig. 1e–h and Extended Data Fig. 4).

The loss of facilitation in Syt7 knockouts cannot be accounted for by slowed recovery from depression reported with Syt7 deletion<sup>19</sup>, because recovery from depression is too slow to influence rapid facilitation strongly, nor can it produce the large increase in release associated with facilitation. There are several possible explanations for the loss of facilitation in knockouts: (1) the presynaptic  $\text{Ca}^{2+}$  signal that induces facilitation could be altered, (2) the probability of release ( $p$ ) for synaptic vesicles could be increased, which by promoting vesicle depletion would indirectly reduce facilitation, or (3) the mechanism for facilitation could be disrupted directly. We assessed these possibilities at the CA3–CA1 synapse.

Action-potential-evoked increases in presynaptic  $\text{Ca}^{2+}$  consist of a large, brief localized  $\text{Ca}^{2+}$  signal that activates the low-affinity  $\text{Ca}^{2+}$  sensor synaptotagmin 1 to trigger neurotransmitter release<sup>23</sup>, and a small residual  $\text{Ca}^{2+}$  signal ( $\text{Ca}_{\text{res}}$ ) that persists for tens of milliseconds and has been implicated in facilitation<sup>2</sup>. It is difficult to measure local  $\text{Ca}^{2+}$  signals that trigger release, but  $\text{Ca}_{\text{res}}$  is readily measured. We used a low-affinity  $\text{Ca}^{2+}$  indicator to measure the time course of  $\text{Ca}_{\text{res}}$  in CA3 presynaptic terminals, because facilitation can be attenuated by the accelerated decay of  $\text{Ca}_{\text{res}}$  (ref. 4).  $\text{Ca}_{\text{res}}$  decayed similarly in wild-type and Syt7-knockout animals (Fig. 2a), indicating that the loss of facilitation in knockout mice is not a consequence of accelerated  $\text{Ca}_{\text{res}}$  decay. We also used  $\text{Ca}_{\text{res}}$  as a measure of  $\text{Ca}_{\text{influx}}$  to determine whether there are use-dependent changes in  $\text{Ca}^{2+}$  entry. However, each of two closely spaced stimuli evoked the same incremental increase in  $\text{Ca}_{\text{res}}$  in both wild types and knockouts (Fig. 2b), indicating that use-dependent changes in total  $\text{Ca}_{\text{influx}}$  cannot account for facilitation. This suggests that if changes in  $\text{Ca}_{\text{influx}}$  contribute to facilitation at this synapse, they must be restricted to the small subset of presynaptic calcium channels that evoke neurotransmitter release. We repeated the experiment using a high-affinity  $\text{Ca}^{2+}$  indicator, in which the degree of saturation during paired stimuli can be used to measure the magnitude of  $\text{Ca}_{\text{res}}$  evoked by the first stimulus (see Methods). We conclude that  $\text{Ca}_{\text{influx}}$  evoked by the first stimulus is the same in wild-type and knockout animals (Fig. 2c).

We further explored the role of  $\text{Ca}^{2+}$  in facilitation by examining the  $\text{Ca}^{2+}$ -dependence of excitatory postsynaptic currents (EPSCs)

<sup>1</sup>Department of Neurobiology, Harvard Medical School, 220 Longwood Avenue, Boston, Massachusetts 02115, USA.



**Figure 1 | Facilitation is absent in Syt7 knockout mice.**

**a–d**, Representative traces (top) and average paired-pulse ratio (PPR) at different interstimulus intervals ( $\Delta t$ ) (bottom) recorded in slices prepared from wild-type (WT; black) and Syt7-knockout (KO; red) animals. Postsynaptic responses were recorded using whole-cell voltage clamp from hippocampal CA1 pyramidal cells (**a**), and thalamic relay cells (**b**). fEPSPs were recorded from hippocampal-mossy-fibre to CA3 synapses (**c**), and lateral-perforant-path synapses in the dentate gyrus (**d**).

Vertical scale bars, 100 pA (**a, b**) and 100  $\mu$ V (**c, d**). **e–h**, Synaptic responses to 20-Hz trains from the same preparations as **a–d** (top), normalized amplitudes during 20-Hz trains (middle), and normalized responses to the tenth stimulus as a function of stimulus frequency (bottom). Peak PPR was significantly different for wild-type and Syt7-knockout mice at all synapses, as was response<sub>10</sub>/response<sub>1</sub> for 5–50-Hz trains ( $P < 0.01$ , Student's *t*-test). Data represent mean  $\pm$  s.e.m. Number of experiments is shown in Extended Data Table 1.

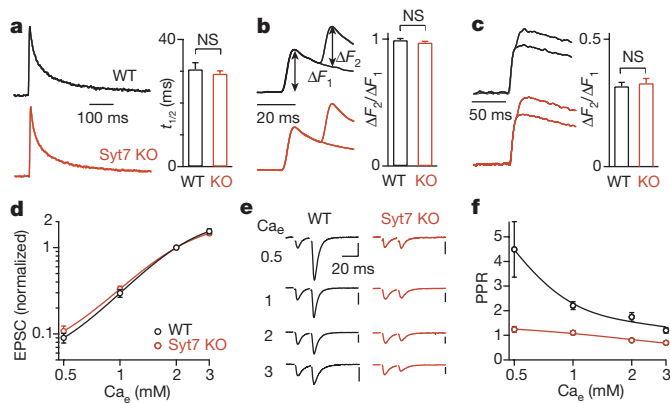
and facilitation. Raising extracellular  $\text{Ca}^{2+}$  leads to a steep increase in EPSC amplitude (Fig. 2d) but a decrease in facilitation (Fig. 2e, black traces), even though high extracellular  $\text{Ca}^{2+}$  should increase the  $\text{Ca}_{\text{res}}$  available to evoke facilitation. This paradox is resolved by realizing that increased  $\text{Ca}^{2+}$  influx increases  $p$ , which depletes presynaptic vesicles, saturates release and limits the extent of facilitation. The  $\text{Ca}^{2+}$ -dependence of EPSC amplitudes was unaffected in knockout animals (Fig. 2d), but facilitation was absent for all values of external  $\text{Ca}^{2+}$  (Fig. 2e, f). Meanwhile, there was no difference in basal release properties measured by the rate of spontaneous EPSCs (Extended Data Fig. 5). These findings suggest that the loss of facilitation in knockouts is not a consequence of higher initial  $p$ , because facilitation was absent even when the initial  $p$  was strongly attenuated by reducing external  $\text{Ca}^{2+}$ .

To test further whether initial  $p$  is increased in Syt7 knockouts, we measured how field excitatory postsynaptic potentials (fEPSPs) scaled with stimulus intensity<sup>24</sup> (Fig. 3a). The slope of the fEPSP versus presynaptic volley gives a relative measure of  $p$  (see Methods), which was unchanged in knockouts (Fig. 3b). Moreover, the fEPSP to presynaptic volley ratio changed steeply with extracellular  $\text{Ca}^{2+}$ , showing that this method is sensitive to  $p$  (Fig. 3c, d). We also assessed  $p$  using pharmacological blockade of synaptically activated NMDARs (*N*-methyl-D-aspartate receptors) by the use-dependent blocker MK801 (ref. 25; Fig. 3e–g). This approach is widely used to detect changes in  $p$ : an increase in  $p$  leads to more glutamate release, and more activation and rapid blockade of NMDARs, while a decrease in  $p$  leads to a slower blockade (Extended Data Fig. 6). The rate of blockade of NMDAR-mediated fEPSPs (NMDAR-fEPSPs) was unaffected by Syt7 deletion (Fig. 3e), indicating similar initial  $p$ . However, when we evoked NMDAR-fEPSPs with trains of three stimuli<sup>25</sup>, amplitudes decayed more rapidly in wild types (Fig. 3f, g), suggesting that Syt7 is required to increase  $p$  for the second and third stimuli. Thus, initial  $p$  and presynaptic  $\text{Ca}^{2+}$

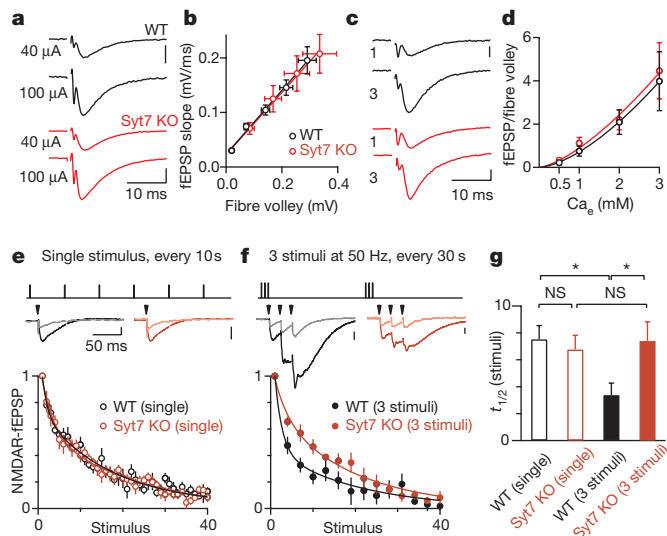
signalling are unaffected by Syt7 deletion, but knockouts lack the use-dependent increase in  $p$  that underlies facilitation. This suggests that the mechanism underlying facilitation is directly impaired by Syt7 deletion.

Syt7 is implicated in neuroendocrine release<sup>16</sup>, insulin secretion<sup>26</sup> and exocytosis of lysosomes<sup>27</sup>, which could all indirectly influence synaptic transmission in global Syt7 knockouts. Therefore, to determine whether Syt7 controls facilitation by acting in presynaptic neurons in a cell-autonomous manner, we tested whether viral expression of Syt7 in CA3 pyramidal cells of Syt7 knockouts rescued facilitation. This approach is complicated by our inability to virally transduce all CA3 pyramidal cells, which prohibits the use of extracellular stimulation that would activate some presynaptic cells that express Syt7 and others that do not. We overcame this problem with an adeno-associated virus (AAV) that drove bicistronic expression of both channelrhodopsin-2 (ChR2) and Syt7, allowing optical stimulation of only those fibres expressing Syt7.

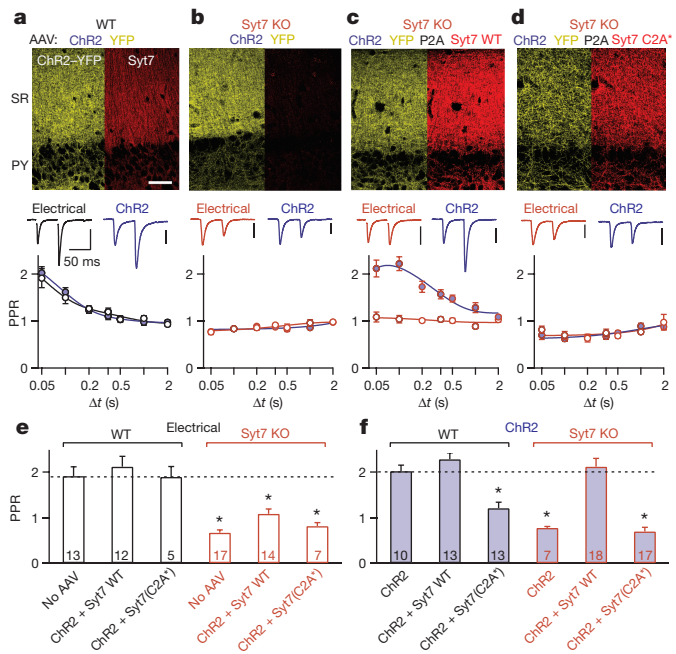
Using conditions we have previously shown allow facilitation to be studied with optogenetic stimulation (see Methods), we confirmed that when ChR2 alone was expressed, optical and electrical stimulation produced similar facilitation in wild types (Fig. 4a, e, f), and similar depression in knockouts (Fig. 4b, e, f). We next used a bicistronic vector to express both ChR2 and wild-type Syt7 in knockout animals. Light-evoked responses exhibited facilitation, whereas electrically evoked responses did not (Fig. 4c, e, f). This suggests that bicistronic expression of ChR2 along with a presynaptic protein of interest offers a powerful new approach to characterize the effect of gene manipulation on presynaptic function within intact neural circuits. When Syt7 was expressed in wild-type animals, the peak facilitation was unaffected (Fig. 4e, f and Extended Data Fig. 7a). Thus, expressing Syt7 in CA3 pyramidal cells rescued facilitation in a cell-autonomous manner, with facilitation restored only at synapses expressing Syt7 and ChR2.



**Figure 2 | Facilitation is altered in Syt7-knockout animals despite similar presynaptic  $\text{Ca}^{2+}$  signals.** **a**, Presynaptic  $\text{Ca}_{\text{res}}$  evoked by a single stimulus recorded from Schaffer collateral fibres loaded with a low-affinity  $\text{Ca}^{2+}$  indicator (left), and  $\text{Ca}_{\text{res}}$  half-decay times (right). NS, not significant. **b**,  $\text{Ca}_{\text{res}}$  signals recorded with low-affinity indicator evoked by one or two stimuli (left). The ratio of the increase in  $\text{Ca}_{\text{res}}$  evoked by the first ( $\Delta F_1$ ) and second ( $\Delta F_2$ ) stimuli (right). **c**,  $\text{Ca}_{\text{res}}$  signals recorded with high-affinity indicator evoked by one or two stimuli. **d**, Average EPSC amplitudes for CA3–CA1 synapses recorded in different external  $\text{Ca}^{2+}$  ( $\text{Ca}_\text{e}$ ) concentrations, normalized to the amplitude in 2 mM  $\text{Ca}_\text{e}$ . **e**, EPSCs recorded in different  $\text{Ca}_\text{e}$ . Vertical scale bars, 50, 100, 200 and 300 pA in 0.5, 1, 2 and 3 mM  $\text{Ca}_\text{e}$ , respectively. **f**, PPR for interstimulus interval of 20 ms recorded in different  $\text{Ca}_\text{e}$ . In 0.5 mM  $\text{Ca}^{2+}$ , the PPR in knockout ( $1.24 \pm 0.12$ ) was not significantly different from 1 ( $P = 0.084$ , Wilcoxon signed rank test). Data represent mean  $\pm$  s.e.m. Number of experiments is shown in Extended Data Table 2.



**Figure 3 | Change in the initial probability of release does not underlie the absence of facilitation in Syt7-knockout mice.** **a**, Extracellular recordings of presynaptic fibre volley and fEPSP evoked by the indicated stimulus intensities. Scale bar, 200  $\mu\text{V}$ . **b**, fEPSP slope plotted against fibre volley amplitude, for 20–100  $\mu\text{A}$  stimulation. **c**, fEPSPs recorded in 1 and 3 mM  $\text{Ca}_\text{e}$ . Scale bar, 100  $\mu\text{V}$ . **d**, Average ratio of the fEPSP to the fibre volley in different  $\text{Ca}_\text{e}$ . **e**, Top, initial release probability was measured by stimulating Schaffer collaterals every 10 s while recording NMDAR-fEPSPs before and after MK801 bath application. Middle, traces averaged from 10 trials before (dark traces), and trials 10–15 after (light traces) MK801 application. Bottom, average NMDAR-fEPSP amplitudes evoked in the presence of MK801. **f**, Same as in **e** but with three stimuli at 50 Hz every 30 s. First response to trains is shown. **g**, Half-decay times of NMDAR-fEPSP amplitudes in the presence of MK801.  $*P < 0.05$ , one-way analysis of variance (ANOVA) with Tukey's post-hoc test. Data represent mean  $\pm$  s.e.m. Number of experiments shown in Extended Data Table 2.



**Figure 4 | Viral expression of Syt7 restores facilitation at Schaffer collateral synapses.** **a–d**, Top, fluorescence images of yellow fluorescent protein (YFP)-tagged ChR2 and Syt7 immunostaining in the CA1 region after AAV injection into CA3 to express the indicated proteins in wild-type animals (**a**) or Syt7-KO animals (**b–d**). PY, stratum pyramidale; SR, stratum radiatum. Scale bar, 100  $\mu\text{m}$ . Bottom, EPSCs and PPRs for responses evoked electrically (open symbols) and optically (blue symbols). In **a** and **b**, only ChR2–YFP was expressed; in **c**, both ChR2–YFP and wild-type Syt7 were expressed (separated by a porcine teschovirus-1 2A (P2A) cleavage peptide); and in **d**, ChR2–YFP and  $\text{Ca}^{2+}$ -insensitive Syt7 (C2A\*) were expressed. **e**, **f**, Summary of PPRs for 50-ms interstimulus interval. Asterisks denote significant difference from responses evoked electrically in uninjected wild-type animals (**e**), or optically in wild-type animals expressing ChR2 alone (**f**).  $*P < 0.05$ , one-way ANOVA with Tukey's post-hoc test. Data represent mean  $\pm$  s.e.m. Number of experiments is shown on bar graphs.

To determine whether  $\text{Ca}^{2+}$  binding by Syt7 is important for facilitation, we assessed whether facilitation is rescued by Syt7 with a mutated  $\text{Ca}^{2+}$ -insensitive C2A domain (Syt7(C2A\*)). Previous studies established that  $\text{Ca}^{2+}$  binding to the C2A domain of Syt7 is required for Syt7 to mediate asynchronous release<sup>18</sup>. We found that Syt7(C2A\*) did not rescue facilitation in knockouts (Fig. 4d–f). Moreover, in wild-type animals, Syt7(C2A\*) expression strongly attenuated facilitation (Fig. 4e, f and Extended Data Fig. 7b), suggesting that Syt7(C2A\*) competes with native Syt7 to suppress facilitation.

Our results indicate that facilitation requires  $\text{Ca}^{2+}$  binding to the C2A domain of Syt7, and also provide insight into the role of Syt7 in facilitation. We conclude that Syt7 does not produce facilitation by altering the amplitude and time course of  $\text{Ca}_{\text{res}}$  (Fig. 2), by increasing initial  $p$  (Fig. 3), by acting as a  $\text{Ca}^{2+}$  buffer (Extended Data Fig. 8), or through use-dependent increases in the total  $\text{Ca}_{\text{influx}}$  (Extended Data Fig. 1b and Fig. 2). The observation that initial  $p$  is unaltered in Syt7 knockouts indicates that local  $\text{Ca}_{\text{influx}}$  is unaffected for the first stimulus, but it is difficult to rule out the possibility that Syt7 mediates a use-dependent increase in  $\text{Ca}_{\text{influx}}$  through the subset of channels that trigger vesicle fusion. There is, however, no evidence for Syt7 associating with or regulating calcium channels. By contrast, Syt7 is known to interact with Syt1 and can mediate vesicle fusion<sup>16–18</sup>. The most parsimonious explanation is that Syt7 acts as the proposed specialized  $\text{Ca}^{2+}$  sensor to increase  $p$  during facilitation. Facilitated release exhibits rapid kinetics, suggesting that Syt7 somehow increases the probability of Syt1-dependent vesicle fusion. Whether this is through a direct interaction of Syt7 with a fast synaptotagmin isoform such as



Syt1 remains an open question. It is also unclear whether the recently described interaction between Syt7 and calmodulin that promotes vesicle replenishment<sup>19</sup> is similarly required for facilitation. Finally, it is possible that at other synapses facilitation is mediated by additional specialized  $\text{Ca}^{2+}$  sensors, or involves other mechanisms. Further studies are needed to clarify these issues.

Based primarily on theoretical considerations, facilitation is thought to influence both information transfer and network dynamics profoundly. In the hippocampus, the high-pass filtering imposed by facilitating synapses may account for the burst firing in place cells that encode spatial information<sup>28</sup>. In the auditory pathway, facilitation is proposed to counteract short-term depression to maintain linear transmission of rate-coded sound intensity<sup>29</sup>. It has even been suggested that facilitation forms the basis of short-term memory, as facilitating recurrent connections within cortical networks could support the persistent activity states associated with working memory<sup>30</sup>. In future studies, the selective elimination of Syt7 from specific cell types could allow the first direct tests of the effect of facilitation on neural circuits and behaviour.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 4 June; accepted 3 December 2015.**

- Feng, T. P. Studies on the neuromuscular junction. XVIII. The local potentials around n-m junctions induced by single and multiple volleys. *Chin. J. Physiol.* **15**, 367–404 (1940).
- Zucker, R. S. & Regehr, W. G. Short-term synaptic plasticity. *Annu. Rev. Physiol.* **64**, 355–405 (2002).
- Abbott, L. F. & Regehr, W. G. Synaptic computation. *Nature* **431**, 796–803 (2004).
- Atluri, P. P. & Regehr, W. G. Determinants of the time course of facilitation at the granule cell to Purkinje cell synapse. *J. Neurosci.* **16**, 5661–5671 (1996).
- Bertram, R., Sherman, A. & Stanley, E. F. Single-domain/bound calcium hypothesis of transmitter release and facilitation. *J. Neurophysiol.* **75**, 1919–1931 (1996).
- Felmy, F., Neher, E. & Schneggenburger, R. Probing the intracellular calcium sensitivity of transmitter release during synaptic facilitation. *Neuron* **37**, 801–811 (2003).
- Südhof, T. C. A molecular machine for neurotransmitter release: synaptotagmin and beyond. *Nature Med.* **19**, 1227–1231 (2013).
- Matveev, V., Zucker, R. S. & Sherman, A. Facilitation through buffer saturation: Constraints on endogenous buffering properties. *Biophys. J.* **86**, 2691–2709 (2004).
- Blatow, M., Caputi, A., Burnashev, N., Monyer, H. & Rozov, A.  $\text{Ca}^{2+}$  buffer saturation underlies paired pulse facilitation in calbindin-D28k-containing terminals. *Neuron* **38**, 79–88 (2003).
- Mochida, S., Few, A. P., Scheuer, T. & Catterall, W. A. Regulation of presynaptic  $\text{CaV}2.1$  channels by  $\text{Ca}^{2+}$  sensor proteins mediates short-term synaptic plasticity. *Neuron* **57**, 210–216 (2008).
- Sippy, T., Cruz-Martin, A., Jeromin, A. & Schweizer, F. E. Acute changes in short-term plasticity at synapses with elevated levels of neuronal calcium sensor-1. *Nature Neurosci.* **6**, 1031–1038 (2003).
- Tsujimoto, T., Jeromin, A., Saitoh, N., Roder, J. C. & Takahashi, T. Neuronal calcium sensor 1 and activity-dependent facilitation of P/Q-type calcium currents at presynaptic nerve terminals. *Science* **295**, 2276–2279 (2002).
- Müller, M., Felmy, F. & Schneggenburger, R. A limited contribution of  $\text{Ca}^{2+}$  current facilitation to paired-pulse facilitation of transmitter release at the rat calyx of Held. *J. Physiol. (Lond.)* **586**, 5503–5520 (2008).
- Hui, E. et al. Three distinct kinetic groupings of the synaptotagmin family: candidate sensors for rapid and delayed exocytosis. *Proc. Natl Acad. Sci. USA* **102**, 5210–5214 (2005).
- Li, C. et al.  $\text{Ca}^{2+}$ -dependent and -independent activities of neural and non-neural synaptotagmins. *Nature* **375**, 594–599 (1995).
- Sugita, S. et al. Synaptotagmin VII as a plasma membrane  $\text{Ca}^{2+}$  sensor in exocytosis. *Neuron* **30**, 459–473 (2001).
- Wen, H. et al. Distinct roles for two synaptotagmin isoforms in synchronous and asynchronous transmitter release at zebrafish neuromuscular junction. *Proc. Natl Acad. Sci. USA* **107**, 13906–13911 (2010).
- Bacaj, T. et al. Synaptotagmin-1 and synaptotagmin-7 trigger synchronous and asynchronous phases of neurotransmitter release. *Neuron* **80**, 947–959 (2013).
- Liu, H. et al. Synaptotagmin 7 functions as a  $\text{Ca}^{2+}$ -sensor for synaptic vesicle replenishment. *eLife* **3**, e01524 (2014).
- Deschênes, M. & Hu, B. Electrophysiology and pharmacology of the corticothalamic input to lateral thalamic nuclei: an intracellular study in the cat. *Eur. J. Neurosci.* **2**, 140–152 (1990).
- Lomo, T. Potentiation of monosynaptic EPSPs in the perforant path-dentate granule cell synapse. *Exp. Brain Res.* **12**, 46–63 (1971).
- Geiger, J. R. & Jonas, P. Dynamic control of presynaptic  $\text{Ca}^{2+}$  inflow by fast-inactivating  $\text{K}^{+}$  channels in hippocampal mossy fiber boutons. *Neuron* **28**, 927–939 (2000).
- Geppert, M. et al. Synaptotagmin I: a major  $\text{Ca}^{2+}$  sensor for transmitter release at a central synapse. *Cell* **79**, 717–727 (1994).
- Dingledine, R. & Somjen, G. Calcium dependence of synaptic transmission in the hippocampal slice. *Brain Res.* **207**, 218–222 (1981).
- Manabe, T. & Nicoll, R. A. Long-term potentiation: evidence against an increase in transmitter release probability in the CA1 region of the hippocampus. *Science* **265**, 1888–1892 (1994).
- Gustavsson, N. et al. Impaired insulin secretion and glucose intolerance in synaptotagmin-7 null mutant mice. *Proc. Natl Acad. Sci. USA* **105**, 3992–3997 (2008).
- Martinez, I. et al. Synaptotagmin VII regulates  $\text{Ca}^{2+}$ -dependent exocytosis of lysosomes in fibroblasts. *J. Cell Biol.* **148**, 1141–1150 (2000).
- Klyachko, V. A. & Stevens, C. F. Excitatory and feed-forward inhibitory hippocampal synapses work synergistically as an adaptive filter of natural spike trains. *PLoS Biol.* **4**, e207 (2006).
- MacLeod, K. M., Horiuchi, T. K. & Carr, C. E. A role for short-term synaptic facilitation and depression in the processing of intensity information in the auditory brain stem. *J. Neurophysiol.* **97**, 2863–2874 (2007).
- Mongillo, G., Barak, O. & Tsodyks, M. Synaptic theory of working memory. *Science* **319**, 1543–1546 (2008).

**Acknowledgements** We thank P. Kaeser and L. Bickford for help with producing AAVs, B. Sabatini and J. Levasseur for help with plasmids, K. Ennis, M. Ocana and the Neurobiology Imaging Center for help with immunohistochemistry, B. Sabatini, P. Kaeser, D. Fioravante, C. Hull and L. Glickfeld for comments on the manuscript. This work was supported by grants from the National Institutes of Health (NIH; NS032405) and Nancy Lurie Marks Foundation to W.G.R., the Vision Core and NINDS P30 Core Center grant (NS072030) to the Neurobiology Imaging Center at Harvard Medical School, and a Nancy Lurie Marks Fellowship to S.L.J.

**Author Contributions** S.L.J., J.T. and W.G.R. designed experiments. J.E.B. performed stereotaxic surgeries, S.L.J. performed electrophysiology, and J.T. measured  $\text{Ca}^{2+}$  and performed immunohistochemistry. S.L.J. and J.T. produced AAVs and analysed experiments, and S.L.J. and W.G.R. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.G.R. ([wade\\_regehr@hms.harvard.edu](mailto:wade_regehr@hms.harvard.edu)).

## METHODS

**Animals and viruses.** All mice were handled in accordance with NIH guidelines and protocols approved by Harvard Medical School. Syt7 knockout mice<sup>31</sup> (Jackson Laboratory) and wild-type littermates of either sex were used. Statistical tests were not used to predetermine sample size. Blinding and randomization were not performed. AAV2/9-hSyn-hChR2(H134R)-EYFP and its pAAV backbone (Addgene 26973) were obtained from the University of Pennsylvania Vector Core. Complementary DNA encoding the rat Syt7 wild-type  $\alpha$  isoform and C2A\* mutant (D225A, D227A and D233A)<sup>18</sup> were provided by T. Bacaj and T. Sudhof. For rescue experiments involving Syt7 with mutated  $\text{Ca}^{2+}$  binding domains, we used the mutated C2A\* version instead of the C2A\* $\text{C2B}^*$  double mutant, as mutation of both C2 domains leads to lower levels of expression. The P2A cleavage sequence<sup>32</sup> and Syt7 were inserted after the ChR2 carboxy terminus in the pAAV backbone (Genscript). Plasmid-driven expression of ChR2-YFP and Syt7 was confirmed in HEK cells by Syt7 immunostaining and patch-clamp recording of ChR2 photocurrents. AAVs were produced and purified from HEK cells as previously described<sup>33</sup>.

Stereotaxic surgeries were performed as described<sup>34</sup>. Postnatal day (P) 18–30 mice were anaesthetized with ketamine/xylazine/acepromazine (100/10/3 mg kg<sup>-1</sup>) supplemented with 1–4% isoflurane. Viruses were injected through glass capillary needles using a syringe (Hamilton) mounted on a stereotaxic instrument (Kopf). Injection coordinates from lambda were 2.69 mm (rostral), 3 mm (lateral) and 2.8 mm (ventral). One microlitre of virus suspension was delivered at a rate of 0.1  $\mu\text{l min}^{-1}$  using a microsyringe pump (WPI; UMP3) and microsyringe pump controller (WPI; Micro4). The needle was slowly retracted 5–10 min after injection, and the scalp incision was closed with gluture. Post-injection analgesic (buprenorphine, 0.05 mg kg<sup>-1</sup>) was administered subcutaneously for 48 h.

**Acute slice preparation.** P30–P60 animals were euthanized under isoflurane anaesthesia, 14–30 days after AAV injection. Brains were removed and placed in ice-cold solution containing (in mM): 234 sucrose, 25 NaHCO<sub>3</sub>, 11 glucose, 7 MgCl<sub>2</sub>, 2.5 KCl, 1.25 NaH<sub>2</sub>PO<sub>4</sub> and 0.5 CaCl<sub>2</sub>. Then, 270- $\mu\text{m}$ -thick transverse slices (hippocampal recordings) or 250- $\mu\text{m}$ -thick sagittal slices (thalamic recordings) were prepared on a vibratome (Leica, VT1000s), and a cut was made between CA3 and CA1 to prevent recurrent excitation. Slices were transferred for 30 min to 32°C artificial cerebrospinal solution (ACSF) containing (in mM): 125 NaCl, 26 NaHCO<sub>3</sub>, 25 glucose, 2.5 KCl, 2 CaCl<sub>2</sub>, 1.25 NaH<sub>2</sub>PO<sub>4</sub> and 1 MgCl<sub>2</sub>, adjusted to 315 mOsm, and allowed to equilibrate to room temperature for >30 min. Experiments were performed at 33  $\pm$  1°C with flow rates of 2 ml min<sup>-1</sup>.

**Electrophysiology.** For ChR2 stimulation, 160 mW mm<sup>-2</sup> laser pulses (0.2–0.5 ms) from a 100-mW 473 nm laser (OptoEngine, MBL-III) were focused through the  $\times 60$  objective of the microscope (Olympus, BX51WI) to produce a 80- $\mu\text{m}$  diameter spot over the stratum radiatum, >500  $\mu\text{m}$  from the recorded cell to avoid activating ChR2 in presynaptic boutons, which can artificially raise the probability of release and obscure facilitation<sup>34</sup>. Extracellular stimulation was performed with a stimulus isolation unit (WPI, A360) using glass monopolar electrodes (0.5–1 M $\Omega$ ) filled with ACSF. Stimulus electrodes were positioned  $\sim$ 500  $\mu\text{m}$  from the recording electrode in the stratum radiatum (Schaffer collaterals), the internal capsule (corticothalamic), the hilus adjacent to the dentate granule cell layer (mossy fibres), and the outer molecular layer (lateral perforant path). To ensure that mossy fibre responses were not contaminated by associational/commissural inputs the metabotropic glutamate receptor agonist DCG-IV (1  $\mu\text{M}$ ) was applied at the end of experiments to block mossy fibre responses selectively<sup>35</sup>. Data were included only if responses were reduced by more than 80% (average reduction was 88  $\pm$  1% in wild-type and 90  $\pm$  2% in Syt7-knockout mice), and the amplitude of mossy fibre responses was measured after subtracting the response remaining in the presence of DCG-IV. Stimulus trials were repeated at 0.1 Hz (0.033 Hz at mossy fibres to avoid potentiation), and artefacts were deleted for display. Recordings were acquired using an amplifier (Axon Instruments, Multiclamp 700B) controlled by custom software written in IgorPro (provided by Matthew Xu-Friedman, SUNY Buffalo), and low-pass filtered at 2 kHz. Whole-cell recordings were obtained using borosilicate patch pipettes (2–5 M $\Omega$ ) pulled with a horizontal puller (Sutter P-97). The internal recording solution contained (in mM): 150 Cs-gluconate, 3 KCl, 10 HEPES, 0.5 EGTA, 3 MgATP, 0.5 NaGTP, 5 phosphocreatine-Tris and 5 phosphocreatine-Na; pH 7.2. Cells were held at -70 mV, and series resistance was monitored during recordings. fEPSPs were recorded in current-clamp mode with ACSF-filled patch pipettes (0.5–1 M $\Omega$ ). Inhibition was blocked with picrotoxin (50  $\mu\text{M}$ ), and during fEPSP recordings, CPP (2  $\mu\text{M}$ ) and CGP (3  $\mu\text{M}$ ) was added to the bath. Approximately 4–10 trials were conducted for each stimulus frequency, and recordings were averaged over trials. Data in all figures represent the mean  $\pm$  s.e.m. Average responses are displayed with double exponential or

polynomial curves fit in IgorPro. Unless stated otherwise, statistical significance was assessed by unpaired two-tailed Student's *t*-test, or one-way ANOVA followed by Tukey's post-hoc test.

**Probability of release.** To record NMDAR-EPSCs, cells were voltage clamped at +40 mV, and the internal solution contained (in mM): 85 Cs-methanesulfonate, 4 NaCl, 10 HEPES, 0.2 EGTA, 30 BAPTA, 2 MgATP, 0.4 NaGTP, 10 phosphocreatine-Na, 25 TEA, 5 QX-314; pH 7.3. For recording NMDAR-fEPSPs, Mg<sup>2+</sup> was excluded from ACSF to relieve Mg<sup>2+</sup> block of NMDA receptors. Picrotoxin (100  $\mu\text{M}$ ) and NBQX (5  $\mu\text{M}$ ) were added to the bath, and stimulation was conducted at 0.1 Hz (unless otherwise indicated) for 5 min to obtain a baseline response. Stimulation was halted for 10 min while (+)-MK801 (40  $\mu\text{M}$ ) was added and allowed to equilibrate. For experiments involving fEPSPs versus presynaptic volley, the postsynaptic response was measured by the slope of the fEPSP, while the amplitude of the presynaptic volley was used to determine the number of activated fibres. If *p* increases, the same number of activated presynaptic fibres will produce a larger fEPSP. The ratio between fEPSP and volley was determined by line fits to the linear regime of the input–output curve of individual experiments (20–80  $\mu\text{A}$  stimuli).

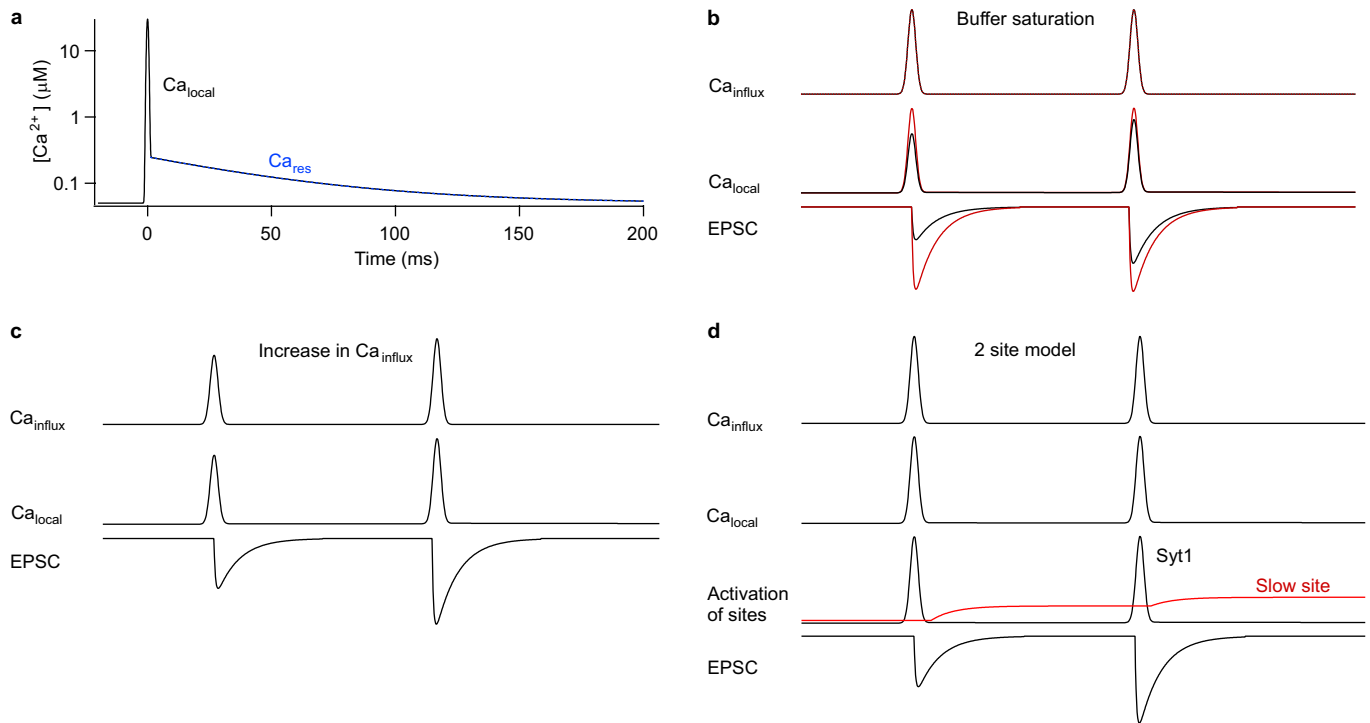
The study of probability of release is complicated because many people use *p* to refer to the probability of release of a vesicle (*p<sub>v</sub>*) and others refer to probability of release from an active zone (*p<sub>synapse</sub>*) that contains *N* vesicles in its readily releasable pool. Thus, an increase in the size of the readily releasable pool for an active zone can increase *p<sub>synapse</sub>* even if *p<sub>v</sub>* is unaltered. Although MK801 blockade<sup>25</sup> and fEPSPs versus presynaptic volley<sup>24</sup> are both widely used methods to detect changes in the probability of release, for both approaches it is conceivable (although unlikely) that increases in *p<sub>v</sub>* could be obscured by a perfectly balanced decrease in the readily releasable pool size. However, the relationship between EPSC amplitude and extracellular Ca<sup>2+</sup> is similar in wild-type and Syt7-knockout animals. This suggests there is no increase in *p<sub>v</sub>*, which would cause this curve to saturate at lower values of Ca<sub>e</sub> for Syt7-knockout animals. Moreover, the large differences in facilitation in wild-type and Syt7-knockout animals were even more pronounced when the probability of release was reduced tenfold by lowering Ca<sub>e</sub> from 2 mM to 0.5 mM, which is incompatible with an increase in *p<sub>v</sub>* obscuring facilitation by depleting vesicles.

**Ca<sup>2+</sup> measurements.** Ca<sup>2+</sup> was measured as described previously<sup>4</sup>. In brief, CA3 fibres were labelled for 3 min using an ACSF-filled pipette containing either magnesium green AM or fura-2 AM (240  $\mu\text{M}$ ) and 1% fast green, placed into the border of the CA3–CA1 field. A vacuum pipette placed above the loading site removed excess indicator. Slices were incubated for at least 1 h and imaging was performed in stratum radiatum of CA1 at least 500  $\mu\text{m}$  from the injection site using a 60 $\times$  objective and custom-built photodiode. Excitation was achieved using a tungsten (magnesium green) or xenon (fura-2) lamp. Schaffer collaterals were stimulated using a glass electrode placed at least 300  $\mu\text{m}$  from the imaging site. To prevent recurrent excitation, experiments were performed in the presence of NBQX (10  $\mu\text{M}$ ), CPP (2  $\mu\text{M}$ ) and picrotoxin (50  $\mu\text{M}$ ).

Magnesium green is a low-affinity calcium indicator<sup>36</sup> ( $K_D = 7 \mu\text{M}$ ) that provides an approximately linear measure of Ca<sub>res</sub> (ref. 37). As such it is well suited to measuring the time course of presynaptic Ca<sub>res</sub> (Fig. 2a) and detecting changes in Ca<sub>influx</sub> during successive stimulations (Fig. 2b). However, with the bulk loading approach the size of the fluorescence change is proportional to the number of stimulated fibres, so the absolute Ca<sub>res</sub> signal is not readily quantified with magnesium green. By contrast, fura-2 has a high affinity for calcium<sup>38,39</sup> ( $K_D = 131 \text{ nM}$ ) so it provides a saturating sublinear response to increases in Ca<sub>res</sub> (refs 40–42). This can be used to test for changes in the absolute size of Ca<sub>influx</sub> because a change in the Ca<sub>influx</sub> per stimulus would change the ratio between the fluorescence change produced by the first and second stimuli.

**Immunohistochemistry.** Two to four weeks after AAV injection, mice were anaesthetized with ketamine and transcardially perfused with 4% paraformaldehyde (PFA) in PBS. The brain was removed and post-fixed for 24 h. Slices (50  $\mu\text{m}$  thick) were permeabilized (PBS plus 0.4% Triton X-100) for 30 min and then prepared in blocking solution (PBS plus 0.2% Triton X-100 and 2% normal goat serum; PBST) for 30 min at room temperature. Slices were incubated overnight at 4°C in PBST with primary antibodies (anti-Syt7 (Synaptic Systems, 105173), 1  $\mu\text{g ml}^{-1}$ ; 1:200, targeting amino acids 46–133 of Syt7 $\alpha$ , anti-vGlut1 (Synaptic Systems, 135304), 1  $\mu\text{g ml}^{-1}$ ; 1:500, and anti-calbindin-D28k (Sigma Aldrich, C9848), 1  $\mu\text{g ml}^{-1}$ ; 1:500), followed by incubation with secondary antibodies in PBST for 2 h at room temperature. For both wild-type and Syt7-knockout mice, images from each brain region were acquired on a laser scanning confocal (Olympus, FluoView1200) using the same laser/microscope settings and processed in ImageJ identically.

31. Chakrabarti, S. *et al.* Impaired membrane resealing and autoimmune myositis in synaptotagmin VII-deficient mice. *J. Cell Biol.* **162**, 543–549 (2003).
32. Kim, J. H. *et al.* High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS ONE* **6**, e18556 (2011).
33. Zolotukhin, S. *et al.* Recombinant adeno-associated virus purification using novel methods improves infectious titer and yield. *Gene Ther.* **6**, 973–985 (1999).
34. Jackman, S. L., Beneduce, B. M., Drew, I. R. & Regehr, W. G. Achieving high-frequency optical control of synaptic transmission. *J. Neurosci.* **34**, 7704–7714 (2014).
35. Kamiya, H., Shinozaki, H. & Yamamoto, C. Activation of metabotropic glutamate receptor type 2/3 suppresses transmission at rat hippocampal mossy fibre synapses. *J. Physiol. (Lond.)* **493**, 447–455 (1996).
36. Zhao, M., Hollingworth, S. & Baylor, S. M. Properties of tri- and tetracarboxylate  $\text{Ca}^{2+}$  indicators in frog skeletal muscle fibers. *Biophys. J.* **70**, 896–916 (1996).
37. Kreitzer, A. C. & Regehr, W. G. Modulation of transmission during trains at a cerebellar synapse. *J. Neurosci.* **20**, 1348–1357 (2000).
38. Brenowitz, S. D. & Regehr, W. G. Calcium dependence of retrograde inhibition by endocannabinoids at synapses onto Purkinje cells. *J. Neurosci.* **23**, 6373–6384 (2003).
39. Grynkiewicz, G., Poenie, M. & Tsien, R. Y. A new generation of  $\text{Ca}^{2+}$  indicators with greatly improved fluorescence properties. *J. Biol. Chem.* **260**, 3440–3450 (1985).
40. Sabatini, B. L. & Regehr, W. G. Detecting changes in calcium influx which contribute to synaptic modulation in mammalian brain slice. *Neuropharmacology* **34**, 1453–1467 (1995).
41. Maravall, M., Mainen, Z. F., Sabatini, B. L. & Svoboda, K. Estimating intracellular calcium concentrations and buffering without wavelength ratioing. *Biophys. J.* **78**, 2655–2667 (2000).
42. Sabatini, B. L. & Svoboda, K. Analysis of calcium channels in single spines using optical fluctuation analysis. *Nature* **408**, 589–593 (2000).
43. Regehr, W. G. Short-term presynaptic plasticity. *Cold Spring Harb. Perspect. Biol.* **4**, a005702 (2012).
44. Kaeser, P. S. & Regehr, W. G. Molecular mechanisms for synchronous, asynchronous, and spontaneous neurotransmitter release. *Annu. Rev. Physiol.* **76**, 333–363 (2014).
45. Kamiya, H. & Zucker, R. S. Residual  $\text{Ca}^{2+}$  and short-term synaptic plasticity. *Nature* **371**, 603–606 (1994).
46. Celio, M. R. Calbindin D-28k and parvalbumin in the rat nervous system. *Neuroscience* **35**, 375–475 (1990).

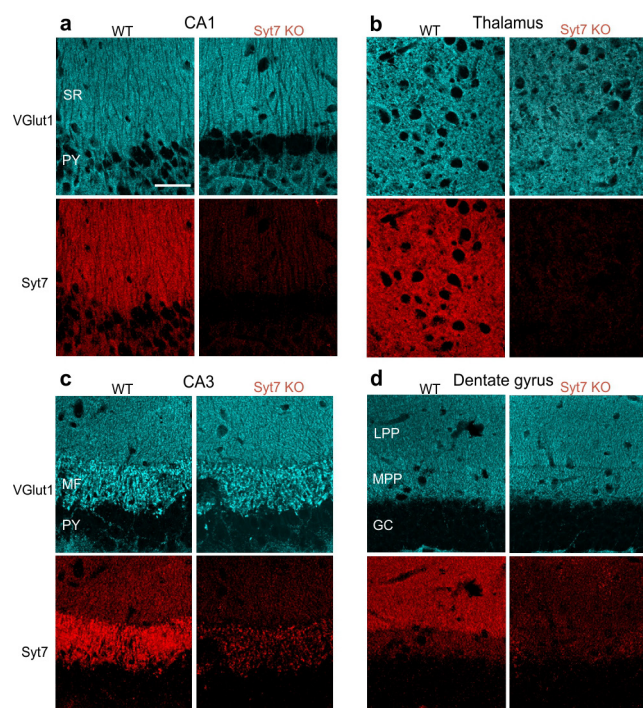


#### Extended Data Figure 1 | Possible mechanisms for synaptic facilitation.

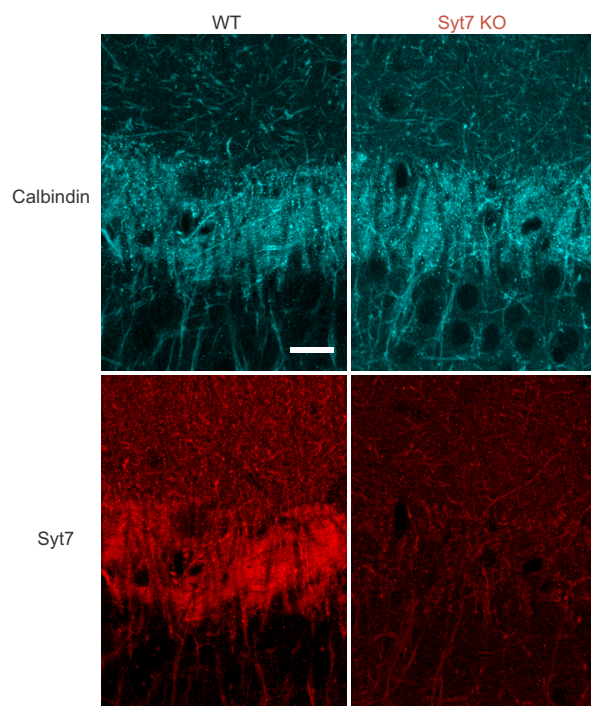
**a–d**, It is established that calcium has an important role in synaptic facilitation, and several mechanisms have been proposed that involve different aspects of calcium signalling<sup>2</sup>. Here we discuss the calcium signals that evoke rapid vesicle fusion, and also those thought to be involved in facilitation (**a**), and three mechanisms of facilitation are presented schematically<sup>43</sup> (**b–d**). **a**, To understand the mechanisms that have been proposed to account for facilitation, it is important to appreciate different aspects of presynaptic calcium signalling. Calcium signals are complex, but can be approximated by two components. An action potential opens calcium channels for less than a millisecond, and near open channels the calcium levels reach tens of micromolar. Release sites near calcium channels experience high local calcium levels ( $Ca_{local}$ ) that are highly dependent on the distance from open calcium channels.  $Ca_{local}$  can be reduced by high concentrations of fast calcium buffers that rapidly bind calcium. In addition, there is a residual calcium signal ( $Ca_{res}$ ) that results from calcium equilibrating within presynaptic terminals, before calcium is gradually removed over tens to hundreds of milliseconds. The amplitude of  $Ca_{res}$  (and also total influx of  $Ca^{2+}$ ,  $Ca_{influx}$ ) is determined by all of the calcium channels that open, not only those that produce  $Ca_{local}$  that drives release, and after initial equilibration  $Ca_{res}$  is roughly uniform throughout the presynaptic bouton. It is generally accepted that fast synaptic transmission is produced by calcium binding to Syt1, Syt2 or Syt9, which have low-affinity binding sites, fast kinetics, and require the binding of multiple calcium ions<sup>7,44</sup>. The time course of release follows the time course of calcium channel opening, but with a brief delay (<1 ms).  $Ca_{res}$  after a single stimulus is much smaller than  $Ca_{local}$ . Typical fluorescence-based approaches to measure calcium readily detect  $Ca_{res}$ ,

but are insensitive to  $Ca_{local}$ , which is too localized and short-lived to measure. Note the y axis is logarithmic to show both  $Ca_{local}$  and  $Ca_{res}$  in **a**, but not in **b–d**. **b**, For one mechanism of facilitation, a fast calcium buffer is present in presynaptic terminals that binds calcium and reduces  $Ca_{local}$ . Stimulation twice in rapid succession results in the same calcium influx for both stimuli. If there is no fast presynaptic buffer, the amplitudes of  $Ca_{local}$  and the EPSCs are the same for both stimuli (red traces). If a fast high-affinity buffer is present (black traces), it reduces the initial  $Ca_{local}$  and reduces the amplitude of the initial EPSC, but if enough calcium enters and binds to the buffer, it reduces its ability to buffer calcium. As a result, the second stimulus produces larger  $Ca_{local}$  than the first, and the EPSC is facilitated. **c**, A second possible mechanism is that more calcium enters for the second stimulus, and as a result there is more neurotransmitter release. This could arise from a spike broadening, or from the modulation of calcium channels. It is possible that influx through all calcium channels in the presynaptic terminal would be increased, in which case both  $Ca_{res}$  and  $Ca_{local}$  would be increased. It is also possible that the only calcium channels that are modulated are the subset that produce  $Ca_{local}$  that triggers release, in which case  $Ca_{res}$  would not be significantly increased. **d**, Finally, it is possible that there is a specialized calcium sensor that produces facilitation that is distinct from Syt1 (refs 2, 4, 45). Previous studies have shown that such a sensor would need to be sensitive to  $Ca_{res}$  based on the observation that facilitation is altered at some synapses by manipulations that affect  $Ca_{res}$  without affecting  $Ca_{local}$ . According to this scheme, release is mediated by Syt1 but calcium binding to a second sensor would increase  $p$ . The sensor is sufficiently slow that it does not influence release evoked by the first stimulus, but it is able to influence release evoked by a second stimulus.

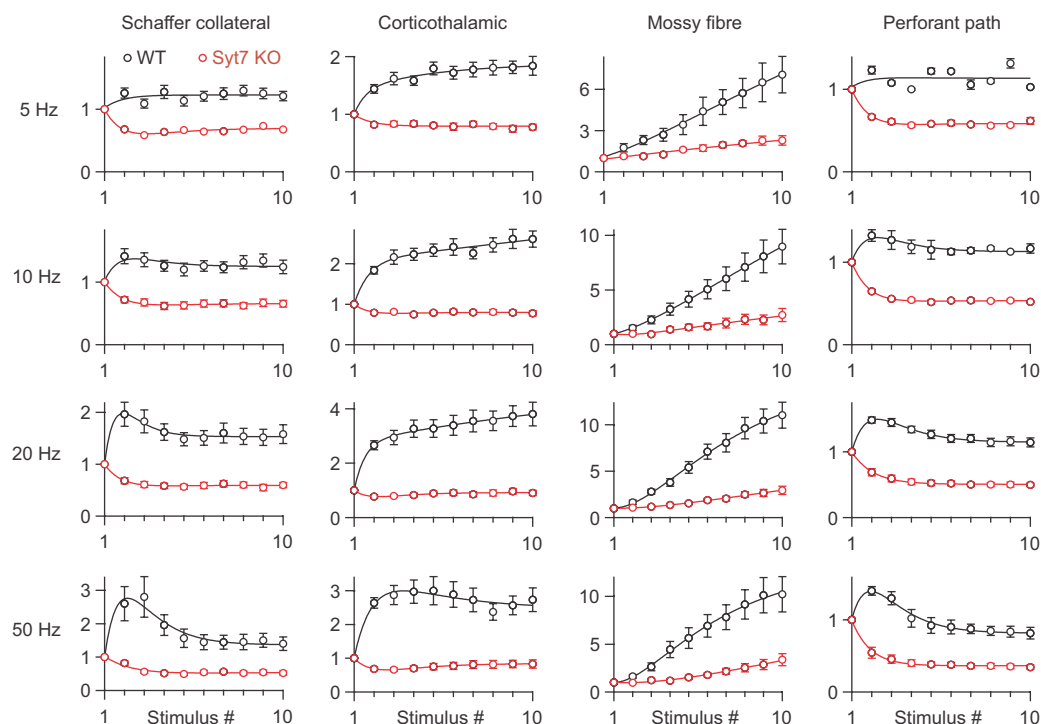




**Extended Data Figure 2 | Immunohistochemistry of Syt7 expression at four different synapses. a–d,** Fluorescent images of immunostaining for vGlut1 (top) and syt7 (bottom) in slices from wild-type and Syt7-knockout animals, showing the stratum radiatum (SR) of hippocampal CA1 region (a), the ventral thalamus (b), mossy fibres (MF) in hippocampal CA3 (c), and the lateral and medial perforant paths (LPP and MPP) in the outer molecular layer of the dentate gyrus (d). Notably, Syt7 expression in wild-type animals was higher in the LPP, where synapses exhibit facilitation, than in the MPP, where synapses exhibit depression. Scale bar, 50  $\mu$ m. The presence of Syt7 labelling in regions containing CA3–CA1 synapses, layer 6 to thalamus synapses, mossy fibres synapses and LPP–granule-cell synapses that are also colabelled with antibodies to the presynaptic marker for glutamatergic synapses vGlut1, suggests that Syt7 is located presynaptically at these synapses. It is, however, difficult to obtain sufficient resolution with confocal microscopy in brain slices to unambiguously establish that Syt7 is located presynaptically at these synapses. Importantly, the Allen Brain atlas (<http://www.brain-map.org>) suggests that the presynaptic cells for these synapses contain messenger RNA for *Syt7*. Lastly, immunoelectron microscopy revealed selective staining of presynaptic boutons in the CA1 region of the hippocampus<sup>16</sup>.

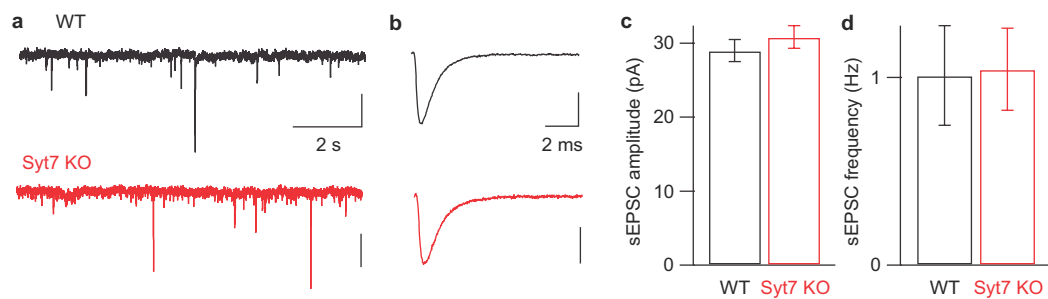


**Extended Data Figure 3 | Immunohistochemistry of Syt7 and calbindin expression at mossy fibre synapses.** Fluorescent images of immunostaining for calbindin-D28k, which predominantly labels mossy fibres in the CA3 region of the hippocampus<sup>9,46</sup> (top) and Syt7 (bottom) in slices from wild-type and Syt7-knockout animals. Colocalization of Syt7 and calbindin staining in wild-type animals provides further support for the expression of Syt7 in mossy fibre terminals. Scale bar, 20  $\mu$ m.



**Extended Data Figure 4 | Loss of facilitation in Syt7-knockout animals at multiple frequencies.** Average normalized synaptic responses evoked by extracellular stimulation with trains at frequencies from 5 to 50 Hz at four synapses in slices from wild-type and Syt7-knockout animals. Enhancement during trains was eliminated for all synapses other than mossy fibre synapses, where significant enhancement was present by the fifth stimulus for 5 Hz and 10 Hz, the third stimulus for 20 Hz, and the

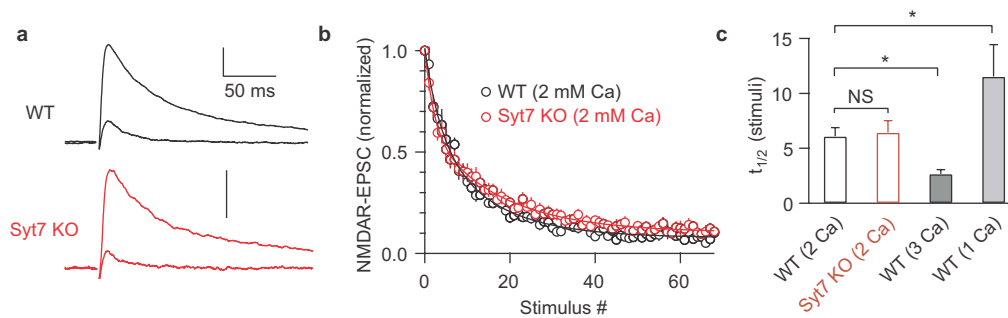
sixth stimulus for 50 Hz (compared to 1 by a Wilcoxon signed rank test,  $P < 0.05$ ). This indicates that another form of synaptic enhancement gradually builds during repetitive activation and is consistent with a specialized form of synaptic enhancement that has been described at mossy fibre synapses in which spike broadening gradually builds during repetitive activation and leads to increased calcium influx. The numbers of experiments are shown in Extended Data Table 1.



**Extended Data Figure 5 | Spontaneous release is similar in wild-type and Syt7-knockout animals.** **a**, Representative spontaneous EPSCs (sEPSCs) recorded from voltage-clamped hippocampal CA1 cells in wild-type (black) and knockout (red) animals. Vertical scale bars, 20 pA.

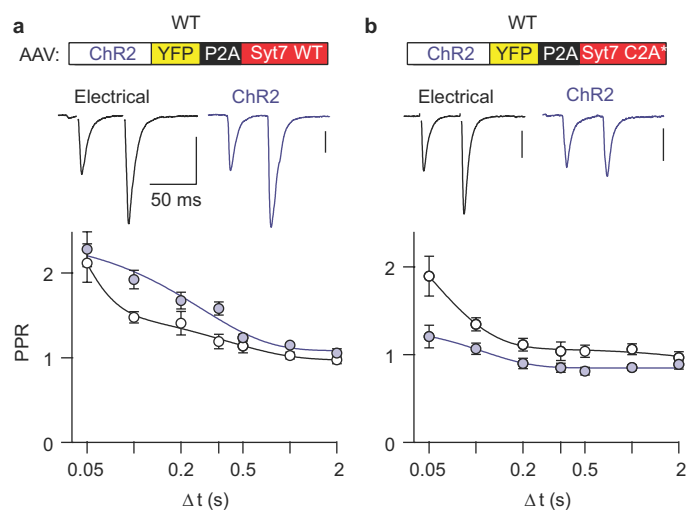
**b**, Representative sEPSCs, averaged from >50 events recorded in wild-type and knockout animals. Vertical scale bars, 10 pA. **c**, **d**, Average sEPSC amplitude (**c**) and frequency (**d**) in wild-type ( $n = 16$ ) and Syt7-knockout animals ( $n = 18$ ).



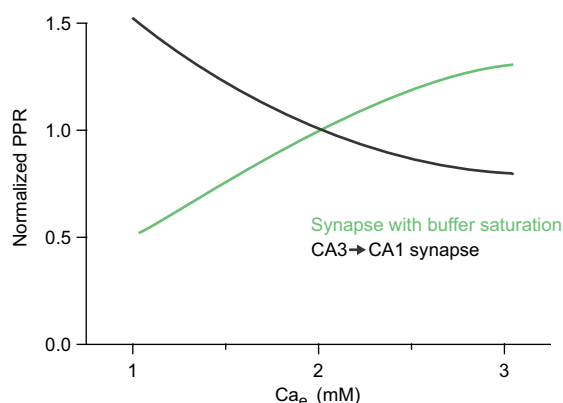


**Extended Data Figure 6 | MK801 blockade of NMDAR-mediated EPSCs reveals similar initial release probability in wild-type and knockout synapses.** **a**, Representative NMDAR-EPSCs recorded in wild-type and knockout animals before the application of MK801 (average of 10 traces) and after stimulation in the presence of MK801 (average response

of fifteenth to twentieth stimuli). Vertical scale bars, 100 pA. **b**, Average NMDAR-EPSCs recorded in the presence of MK801, normalized to the first stimulus. **c**, Half-decay times of NMDAR-EPSC amplitudes. \* $P < 0.05$ , one-way ANOVA with Tukey's post-hoc test. Data represent mean  $\pm$  s.e.m. The number of experiments is shown in Extended Data Table 2.



**Extended Data Figure 7 | Effect of virally expressed Syt7 wild-type and Syt7(C2A\*) in wild-type animals. a, b,** Top, AAV was injected into the hippocampal CA3 region in wild-type animals to express ChR2 and either wild-type Syt7 (**a**) or Syt7(C2A\*) (**b**). Bottom, representative EPSCs and average paired-pulse ratios for responses evoked electrically and optically in wild-type slices with AAV-driven expression of wild-type Syt7 (electrical,  $n = 12$ ; optical,  $n = 13$ ) (**a**) and Syt7(C2A\*) (electrical,  $n = 5$ ; optical,  $n = 13$ ) (**b**). Vertical scale bars, 100 pA.



**Extended Data Figure 8 | Evidence suggests that Syt7 does not produce facilitation by acting as a local calcium buffer at the CA3–CA1 synapse.**

This graph illustrates the general relationship between PPR and external calcium for synapses in which buffer saturation produces facilitation (green) and for facilitation observed at the CA3–CA1 synapse and many other synapses (black)<sup>9</sup>. It has been shown previously that the for buffer saturation mechanism (Extended Data Fig. 1b) the amplitude of facilitation is reduced when  $Ca_{influx}$  is reduced by lowering external calcium<sup>9</sup>. This can be understood by considering that this form of facilitation is thought to require sufficient  $Ca_{influx}$  to saturate the endogenous buffer, and thereby reduce its ability to buffer calcium for subsequent stimuli. If  $Ca_{influx}$  is low, then there is insufficient calcium entry to bind very much of the endogenous buffer, and little facilitation would result. In addition, as shown in Extended Data Fig. 1, for a calcium buffer to produce facilitation it would need to buffer calcium sufficiently that it would reduce initial  $p$ . We have shown, however, that  $p$  is unaltered in Syt7 knockouts. This is perhaps not surprising in light of the fact that Syt7 is thought to be located on the plasma membrane, and in cases where this type of facilitation has been observed it is associated with high concentrations of a fast cytosolic buffer<sup>9</sup>.

**Extended Data Table 1 | Number of electrophysiological recordings from wild-type and Syt7-knockout animals**

Figure	Synapse	Experiment	Genotype	# of Recordings	# of Animals
Figure 1a	Schaffer collateral	Paired-pulse	WT	13	6
			KO	17	5
Figure 1b	Corticothalamic	Paired-pulse	WT	23	8
			KO	23	5
Figure 1c	Hippocampal mossy fibre	Paired-pulse	WT	10	6
			KO	8	4
Figure 1d	Lateral perforant path	Paired-pulse	WT	6	3
			KO	13	3
Figure 1e, Extended	Schaffer collateral	Trains (2-50 Hz)	WT	14	6
Data Figure 4		Trains (2-50 Hz)	KO	17	5
Figure 1f, Extended	Corticothalamic	Trains (2-50 Hz)	WT	16	8
Data Figure 4		Train (2 Hz)	KO	5	2
		Trains (5-50 Hz)	KO	12	3
Figure 1g, Extended	Hippocampal mossy fibre	Train (2 Hz)	WT	3	2
Data Figure 4		Train (5 Hz)	WT	7	4
		Train (10-50 Hz)	WT	10	5
		Train (2-5 Hz)	KO	3	2
		Train (10-50 Hz)	KO	8	4
Figure 1h, Extended	Lateral perforant path	Train (2-10 Hz)	WT	3	2
Data Figure 4		Train (20-50 Hz)	WT	6	3
		Train (2-5 Hz)	KO	5	2
		Train (10 Hz)	KO	10	3
		Train (20-50 Hz)	KO	13	3



**Extended Data Table 2 | Number of experiments related to the  $\text{Ca}^{2+}$ -dependence of probability of release**

Figure	Experiment	Condition	Genotype	# of Recordings	# of Animals
Figure 2a,b	Presynaptic $\text{Ca}^{2+}$ imaging	Magnesium Green	WT	11	2
			KO	10	2
Figure 2c	Presynaptic $\text{Ca}^{2+}$ imaging	Fura-2	WT	14	3
			KO	10	2
Figure 2d-f	$\text{Ca}^{2+}$ -dependence of CA3-CA1 EPSC	0.5 mM Ca	WT	12	5
		1 mM Ca	WT	9	4
		2 mM Ca *	WT	15	6
		3 mM Ca	WT	6	2
		0.5 mM Ca	KO	8	4
		1 mM Ca	KO	7	6
		2 mM Ca *	KO	10	8
		3 mM Ca	KO	4	2
		20-100 uA stimulation	WT	44	11
			KO	25	8
Figure 3a,b	fEPSP vs. fibre volley	0.5 mM Ca	WT	4	2
		1 mM Ca	WT	11	5
		2 mM Ca *	WT	11	5
		3 mM Ca	WT	8	3
		0.5 mM Ca	KO	4	4
		1 mM Ca	KO	8	4
		2 mM Ca *	KO	9	5
		3 mM Ca	KO	6	3
		2 mM Ca, single stim	WT	6	2
			WT	5	3
Figure 3e-g	MK801 blockade of NMDAR-fEPSP	2 mM Ca, single stim	KO	6	3
		2 mM Ca, triple stim	KO	4	3
		1 mM Ca	WT	14	3
		2 mM Ca	WT	11	4
		3 mM Ca	WT	3	2
		2 mM Ca	KO	9	4
Extended Data Figure 6	MK801 blockade of NMDAR-EPSC	1 mM Ca	WT	14	3
		2 mM Ca	WT	11	4
		3 mM Ca	WT	3	2
		2 mM Ca	KO	9	4

\*To normalize responses in different  $\text{Ca}^{2+}$  concentrations, all  $\text{Ca}^{2+}$ -dependence experiments included recordings in 2 mM  $\text{Ca}^{2+}$  followed by wash in of different  $\text{Ca}^{2+}$  concentrations.

# The *C. elegans* adult neuronal IIS/FOXO transcriptome reveals adult phenotype regulators

Rachel Kaletsky<sup>1\*</sup>, Vanisha Lakhina<sup>1\*</sup>, Rachel Arey<sup>1</sup>, April Williams<sup>1</sup>, Jessica Landis<sup>1</sup>, Jasmine Ashraf<sup>1</sup> & Coleen T. Murphy<sup>1</sup>

**Insulin/insulin-like growth factor signalling (IIS) is a critical regulator of an organism's most important biological decisions from growth, development, and metabolism to reproduction and longevity. It primarily does so through the activity of the DAF-16 transcription factor (forkhead box O (FOXO) homologue), whose global targets were identified in *Caenorhabditis elegans* using whole-worm transcriptional analyses more than a decade ago<sup>1</sup>. IIS and FOXO also regulate important neuronal and adult behavioural phenotypes, such as the maintenance of memory<sup>2</sup> and axon regeneration<sup>3</sup> with age, in both mammals<sup>4</sup> and *C. elegans*, but the neuron-specific IIS/FOXO targets that regulate these functions are still unknown. By isolating adult *C. elegans* neurons for transcriptional profiling, we identified both the wild-type and IIS/FOXO mutant adult neuronal transcriptomes for the first time. IIS/FOXO neuron-specific targets are distinct from canonical IIS/FOXO-regulated longevity and metabolism targets, and are required for extended memory in IIS *daf-2* mutants. The activity of the forkhead transcription factor FKH-9 in neurons is required for the ability of *daf-2* mutants to regenerate axons with age, and its activity in non-neuronal tissues is required for the long lifespan of *daf-2* mutants. Together, neuron-specific and canonical IIS/FOXO-regulated targets enable the coordinated extension of neuronal activities, metabolism, and longevity under low-insulin signalling conditions.**

The *C. elegans* IIS pathway acts both cell autonomously and non-autonomously to control longevity, growth, dauer formation, metabolism, and reproduction<sup>5–7</sup> through its regulation of the nuclear localization and transcriptional activation of DAF-16 (also known as FOXO). The canonical IIS/FOXO gene set, which identified primarily intestinal and hypodermal targets (Extended Data Fig. 1a, b)<sup>1,8,9</sup>, has been instructive in our understanding of how insulin signalling regulates a diverse range of activities, including metabolism, autophagy, stress resistance, and proteostasis. However, IIS mutants also exhibit *daf-16*-dependent neuronal phenotypes, including extended positive olfactory learning<sup>2</sup>, increased short- and long-term associative memory<sup>2</sup>, increased thermotaxis learning<sup>10</sup>, improved neuronal morphology maintenance<sup>11,12</sup>, and improved axon regeneration<sup>3</sup>. These phenotypes are unlikely to be regulated by the known intestinal and hypodermal IIS/FOXO targets<sup>1,8</sup>. Therefore, to understand how IIS *daf-2* mutant animals extend behavioural functionality, we must identify the neuronal targets of FOXO/DAF-16.

We first profiled the expression of *daf-16*; *daf-2* mutant worms with *daf-16* rescued in specific tissues<sup>6</sup> (Supplementary Table 1). Intestinal *daf-16* rescue correlates best with whole-worm profiles (Extended Data Fig. 1a, c). By contrast, neuronal *daf-16* rescue profiles are anti-correlated with the intestinal DAF-16 and whole-worm profiles (Extended Data Fig. 1a, c). Surprisingly, many genes induced by neuronal DAF-16 rescue are expressed (WormBase) or predicted to be expressed in non-neuronal tissues<sup>13</sup> (Extended Data Fig. 1d), and have non-neuronal functions (for example, collagens<sup>14</sup>; Extended

Data Fig. 1b, e, Supplementary Table 2). Thus, whole-worm transcriptional analyses of neuronally rescued DAF-16 failed to reveal targets that account for *daf-16*-dependent age-related behaviours of *daf-2* mutants. Therefore, we needed to specifically examine transcription in IIS-mutant neurons.

The tough outer cuticle prevents dissociation of adult tissues<sup>15</sup>, thus the wild-type adult neuronal transcriptome has not been described. To solve this problem, we used rapid, chilled chemomechanical disruption followed immediately by fluorescence-activated cell sorting (FACS) to isolate neurons marked with green fluorescent protein (GFP) from wild-type worms, then RNA-sequenced these isolated cells (Fig. 1a–c, Extended Data Fig. 2a–c, f, g, Supplementary Table 3). This method is gentle enough to preserve the integrity of cells and some neurites (Extended Data Fig. 2a), does not involve cell culturing before FACS, in contrast to previous methods<sup>16</sup>, and does not affect transcription (as shown by actinomycin D treatment; Fig. 1b, Extended Data Fig. 2d, e, Supplementary Table 4). Downsampling analysis showed that sufficient sequencing depth was achieved (Extended Data Fig. 2h).

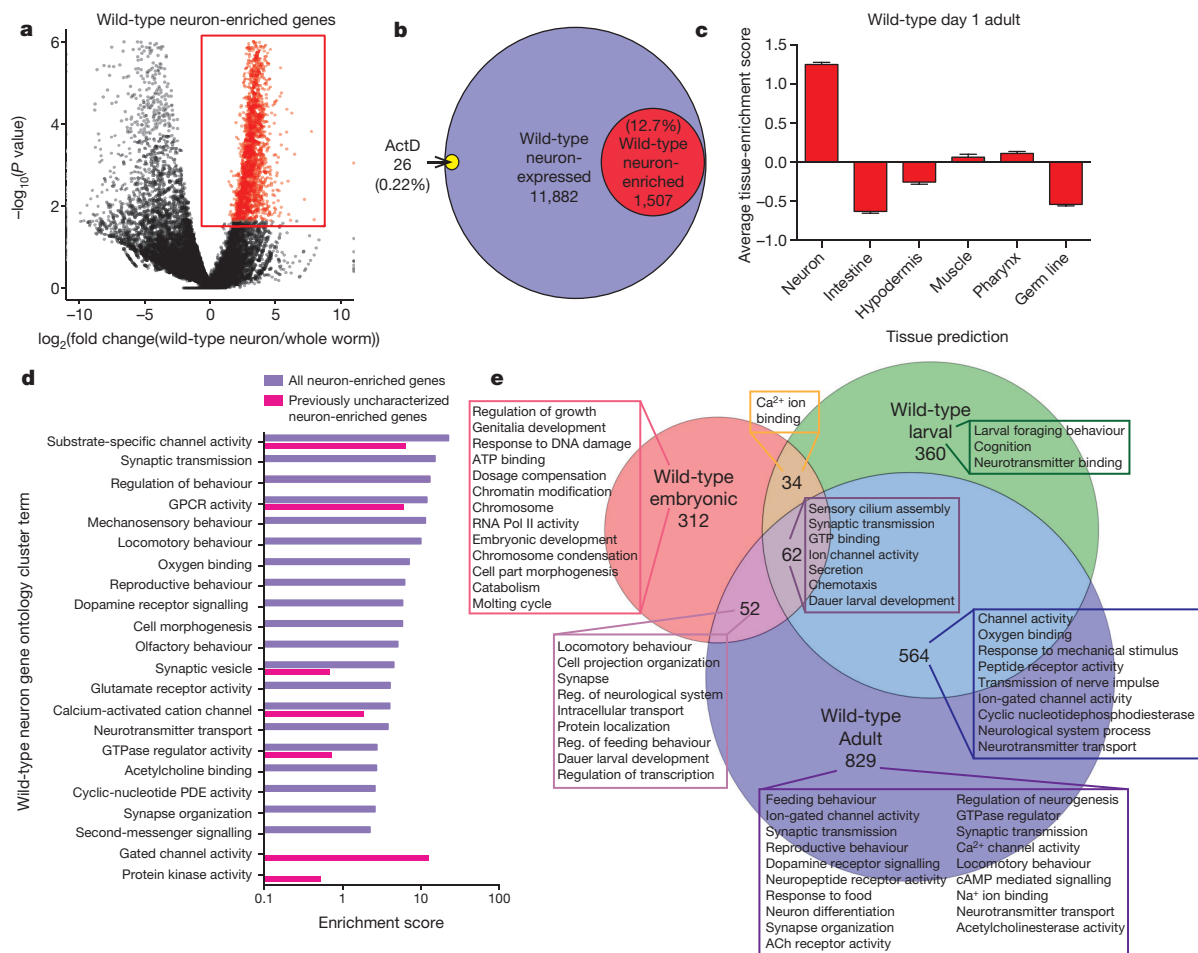
We compared gene expression in isolated wild-type neurons with whole-worm expression to identify genes that are enriched in neurons (Fig. 1a–c). Of the 1,507 'neuron-enriched' genes (false discovery rate (FDR) < 0.1; Supplementary Table 3; Fig. 1a, b), only 4% have previously described expression patterns exclusively in non-neuronal tissues, and 'Neuron' is the only significantly enriched tissue (Fig. 1c, Extended Data Fig. 2f), indicating that the method is highly selective for neuronal transcripts. Gene promoter–GFP tests of previously uncharacterized genes from our neuron-enriched list confirmed neuronal expression, with no bias for particular neuron types (Extended Data Fig. 3a). We also detected genes previously reported to be expressed only in single neurons or small subsets of neurons, including *glr-3* (in the RIA neuron), *ttx-3* (interneuron AIY/AIA) and *npr-14* (neuron AIY) (WormBase).

The wild-type neuron-enriched set includes synaptic machinery, ion channels, neurotransmitters, and signalling components (Supplementary Table 3), as well as >700 previously uncharacterized genes; these genes are predicted to have 'neuronal'-like character and function (Fig. 1d). Comparison of the wild-type embryonic and larval neuronal transcriptomes with the adult neuronal transcriptome at the same FDR revealed a shift in functional categories from developmental processes to neuronal function/behaviour in the adult neuronal transcriptome (Fig. 1e, Extended Data Fig. 3b, c, Supplementary Table 5), suggesting that previous isolation methods<sup>16</sup>, either due to early developmental stage isolation or to re-culturing, biased expression towards developmental genes rather than neuronal/behavioural genes.

To identify adult neuronal IIS/FOXO targets, we sequenced RNA from isolated *daf-2* and *daf-16*; *daf-2* mutant neurons on day 1 of adulthood (Fig. 2a, Extended Data Fig. 4, Supplementary Table 6, 8). The IIS/FOXO neuron-isolated gene set is enriched for neuronal expression: 86% and 92% of the up- and downregulated genes, respectively, are expressed in wild-type neurons. While several top Class I gene

<sup>1</sup>Department of Molecular Biology & LSI Genomics, Princeton University, Princeton, New Jersey 08544, USA.

\*These authors contributed equally to this work



**Figure 1 | Identification of neuronal IIS/FOXO targets requires neuronal isolation.** **a**, Volcano plot of neuron-expressed relative to whole-worm-expressed genes obtained by neuron-specific RNA sequencing of adult wild-type animals.  $N = 3$  biological replicates (wild-type neurons) and 2 biological replicates (whole worm). **b**, Neuron-expressed and -enriched genes are not influenced by cell isolation: treatment with the transcription inhibitor actinomycin D affected only 0.22% of all neuronal genes

targets of DAF-16, including *hil-1*, *sip-1*, *mtl-1*, *nnt-1*, *ins-6*, and *daf-16* itself, were upregulated in both *daf-2* mutant neurons and *daf-2* mutant whole worms (Group B; Fig. 2b), most of the IIS/FOXO neuronally regulated set differs from the canonical whole-worm IIS/FOXOs set<sup>1,8</sup> (Fig. 2b). Specifically, in contrast to the metabolism-dominated functions of canonical whole-worm IIS/FOXO targets<sup>1,8</sup>, the neuronal IIS set gene ontology terms reflect neuron-like functions (Extended Data Fig. 5b): serpentine receptors, G protein-coupled receptors, syntaxin, globins, kinesins, insulins, ion channels, potassium channels, seven-transmembrane receptors, the NPR-1 neuropeptide receptor, and the SER-3 octopamine receptor are upregulated in *daf-2* neurons (Supplementary Table 6). A few genes (*fat-3* and *crh-1*, a CREB homologue) are upregulated in *daf-2* neurons but downregulated in whole *daf-2* animals.

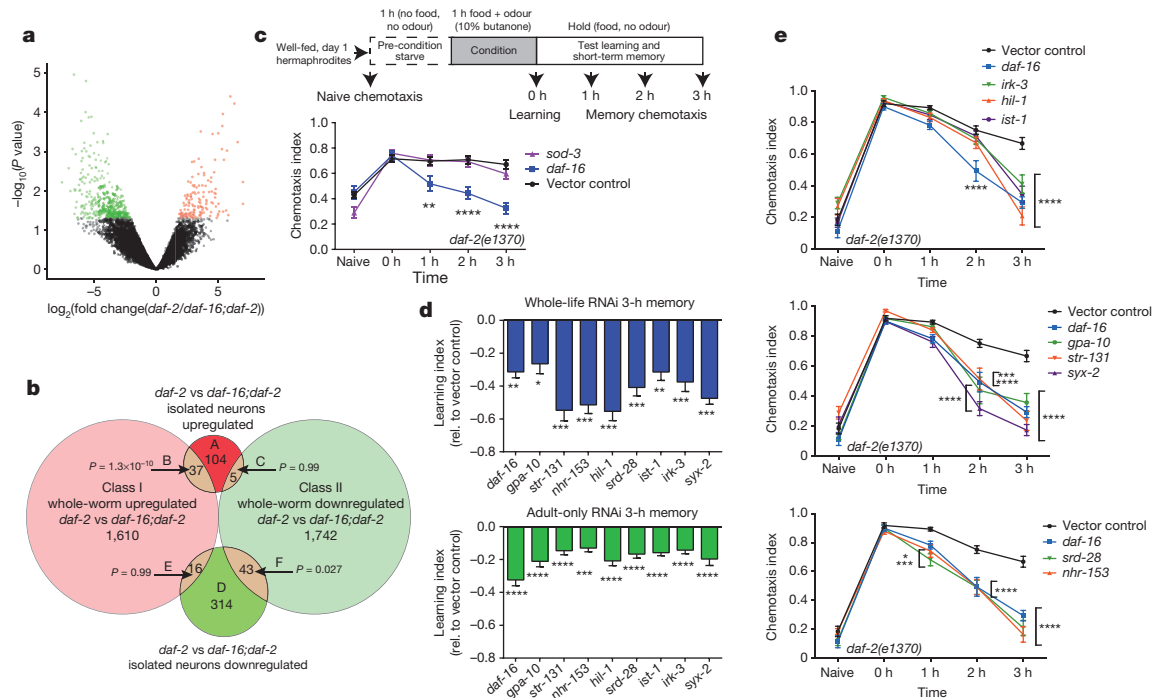
The IIS/FOXO downregulated set includes serpentine receptors, guanylate cyclases, signalling peptides and receptors (neuropeptide-like proteins, FMRF-like peptides and neuropeptides), and the vesicle trafficking G protein *rab-28* (Supplementary Table 6). Expression of the sensory neuron cilia protein IFTA-2, which co-localizes with DAF-2 and whose loss increases lifespan<sup>17</sup>, is downregulated in *daf-2* mutants, consistent with the longevity of *daf-2* and ciliated sensory neuron mutants<sup>18</sup>. Similarly, *sams-1* (S-adenosyl methionine synthetase), which is downregulated under long-lived dietary restriction conditions<sup>19</sup>, and *sma-5* and *dbl-1*, components of TGF-beta pathways

(Supplementary Table 4). **c**, Tissue expression prediction of wild-type adult neuron-enriched genes. Mean  $\pm$  s.e.m. **d**, GO terms highlight the neuronal characteristics of both all and previously uncharacterized neuron-enriched genes. **e**, Embryonic<sup>16</sup>, larval<sup>16</sup> and adult neuron-enriched genes and significant GO terms transition from developmental to neuronal and behavioural functions (Supplementary Table 5); FDR < 10% for all gene sets.

linked with IIS<sup>7,20</sup>, are downregulated, perhaps coordinating the longevity and reproductive output of these pathways.

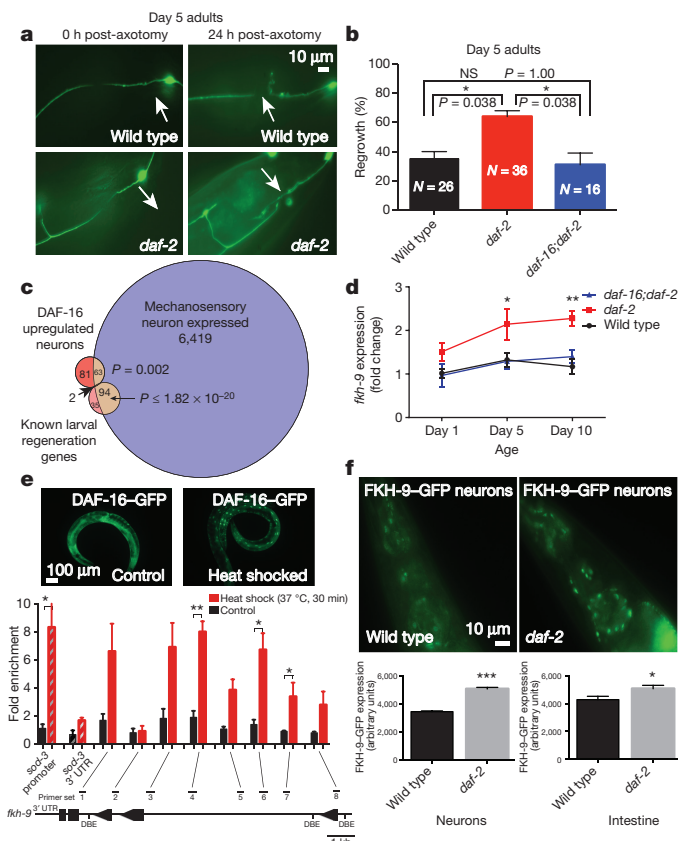
Unlike canonical IIS/FOXO targets<sup>1</sup>, neuronal IIS/FOXO gene promoters are not enriched for the DBE (DAF-16 binding element, GTAAAT/cA), but the overlapping, upregulated (Group B) targets' promoters contain twice as many DBEs (Extended Data Fig. 5a). The overlapping downregulated (Group F) targets are enriched for the PQM-1/DAE motif (CTTATCA, see refs 1, 8; Supplementary Table 7). DAF-16 may regulate neuronal activities indirectly through activation of  $\sim 60$  IIS/FOXO-upregulated transcription factors (Supplementary Table 6).

We next tested the roles of top-scoring genes in *daf-2*-regulated neuronal phenotypes. Long-term and short-term associative memory are both extended in *daf-2* mutants in a *daf-16*-dependent manner<sup>2</sup> (Extended Data Fig. 6). The bZIP transcription factor CREB, which is required for long-term memory in many organisms, including *C. elegans*<sup>2</sup>, is upregulated by IIS/FOXO in neurons (Supplementary Table 6), correlating with the increased long-term memory of *daf-2* mutants<sup>2,21</sup>. However, short-term associative memory (STAM; Fig. 2c) is CREB-independent<sup>2</sup>, and the genes that enable STAM extension in *daf-2* mutants are unknown. While the DAF-16 non-neuronal target *sod-3* had no effect on the extended STAM of *daf-2* mutants (Fig. 2c, Extended Data Fig. 6b–d), knockdown of 8 of the 10 top-ranked, upregulated IIS/FOXO targets significantly decreased the STAM of



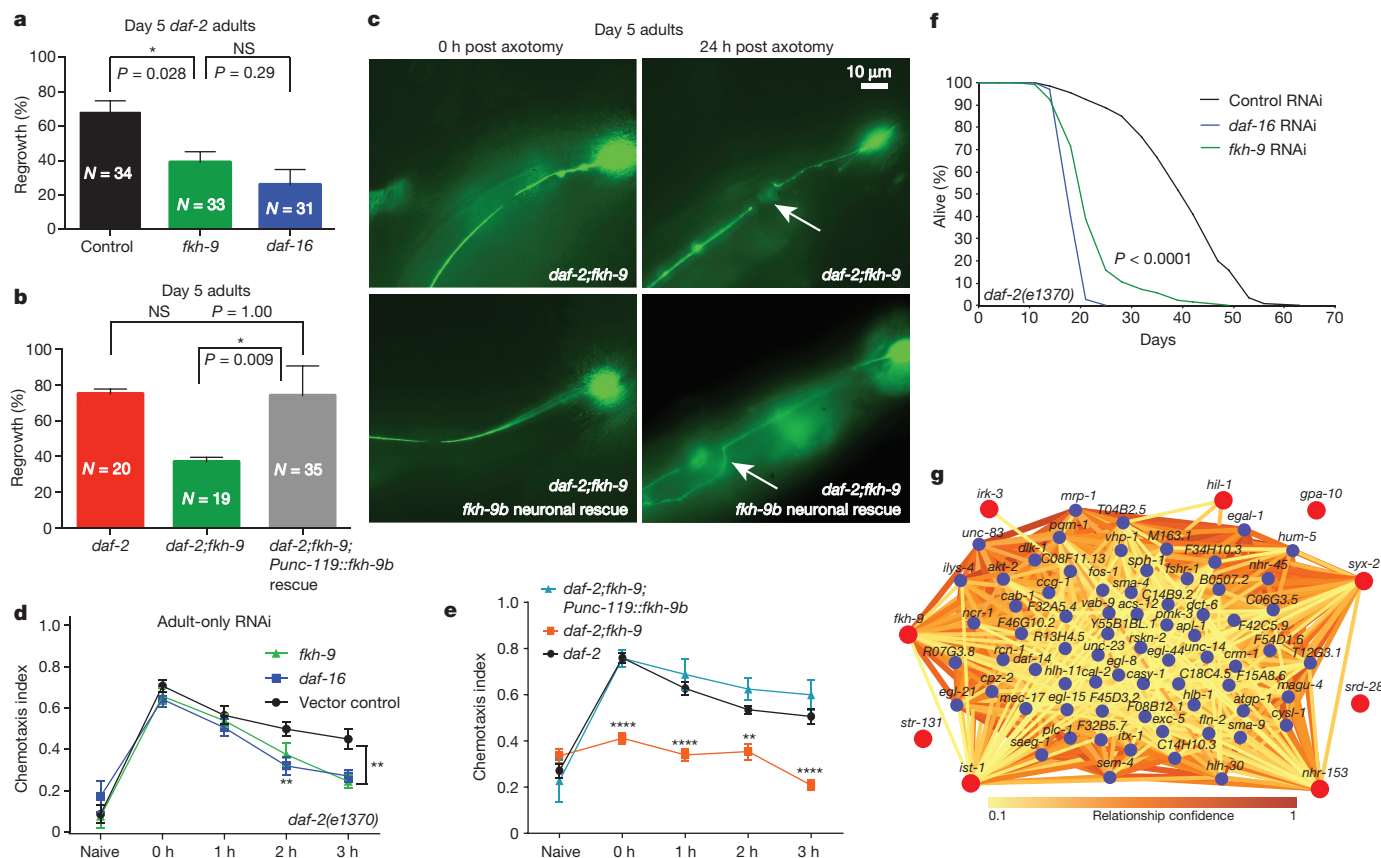
**Figure 2 | RNA-seq transcriptional profile of isolated neurons reveals IIS/FOXO neuronal transcriptome.** **a**, Volcano plot of *daf-2*-regulated, *daf-16*-dependent up- (red) and downregulated (green) neuronal genes ( $P < 0.05$ ,  $N = 4$  biological replicates per strain). **b**, Comparison of whole-worm (Class I)<sup>8</sup> vs neuronal-IIS/FOXO targets.  $P$  values: hypergeometric distributions. **c–e**, Short-term associative memory (STAM) assays. **c**, Schematic of STAM assay and chemotaxis profiles of *daf-2* treated with *sod-3* (c) or neuronal

IIS/FOXO target gene RNAi (d, e). **d**, Learning indices relative to control RNAi at 3 h post-training of *daf-2* treated with adult-only (green) or whole-life (blue) neuronal IIS/FOXO target gene RNAi. Mean  $\pm$  s.e.m.,  $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$ ,  $****P < 0.0001$ , two-way repeated measures ANOVA, Bonferroni post hoc tests. At least 3 biological replicates were performed for all STAM assays.



**Figure 3 | FKH-9 is a direct target of DAF-16 and is expressed in mechanosensory neurons.** **a**, *daf-16* is required for enhanced day 5 axon regeneration in *daf-2* mutants, mean  $\pm$  s.e.m.,  $*P < 0.05$ , Fisher's exact test,  $N = 26$  (wild-type), 36 (*daf-2*) and 16 (*daf-16;daf-2*), 2 biological replicates. **c**, Known larval regeneration genes are significantly enriched in the day 1 adult mechanosensory transcriptome. 63 genes are both DAF-16 targets and expressed in mechanosensory neurons ( $FDR < 5\%$ ; 3 biological replicates). **d**, *fkh-9* messenger RNA levels are higher in aged *daf-2* compared to wild type in a *daf-16*-dependent manner.  $N = 4$  biological replicates, two-way ANOVA, Bonferroni post hoc tests. **e**, Chromatin immunoprecipitation of DAF-16-GFP worms with and without heat shock, which mobilizes DAF-16 into the nucleus. DAF-16 binds to the *sod-3* promoter but not its 3' UTR, and to the *fkh-9* promoter at multiple locations (Extended Data Fig. 8). Fold enrichment relative to wild-type (not expressing DAF-16-GFP) is shown (mean  $\pm$  s.e.m., two-tailed  $t$ -test,  $N = 3$  biological replicates). **f**, Neuronal FKH-9-GFP (*fkh-9p::fkh-9::gfp*) expression in *daf-2* compared to wild type.  $N = 25$  animals. Mean  $\pm$  s.e.m., two-tailed  $t$ -test. **d–f**,  $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$ .





**Figure 4 | FKH-9 is required for improved axon regeneration, short-term associative memory and lifespan in *daf-2* mutants.**

**a**, *fkh-9* knockdown reduces axon regeneration of day 5 *daf-2* mutants, as does *daf-16* knockdown. Mean  $\pm$  s.e.m., \* $P < 0.05$ , Fisher's exact test,  $N = 34$  (control), 33 (*fkh-9*) and 31 (*daf-16*), 4 biological replicates. **b**, **c**, Neuronally-expressed *fkh-9* rescues day 5 axon regeneration in *daf-2;fkh-9* mutants. Mean  $\pm$  s.e.m., \* $P < 0.05$ , Fisher's exact test,  $N = 20$  (*daf-2*), 19 (*daf-2;fkh-9*) and 35 (*daf-2;fkh-9; Punc-119::fkh-9b*), 2 biological replicates. **d**, *fkh-9* is required for enhanced memory in adult-only RNAi-treated *daf-2* mutant worms. **e**, Neuronally-expressed *fkh-9*

rescues extended STAM in *daf-2;fkh-9* mutants with defective learning and memory. Mean  $\pm$  s.e.m., \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$ , two-way repeated measures ANOVA, Bonferroni post hoc tests. **f**, Adult-specific *fkh-9* RNAi treatment reduces *daf-2* mutant lifespan. Median lifespan: control RNAi 42 days, *fkh-9* RNAi 21 days, *daf-16* RNAi 21 days.  $P < 0.0001$  for control RNAi vs *daf-16* RNAi and control vs *fkh-9* RNAi, log-rank test.  $N = 144$  worms per strain. **g**, Integrative Multi-species Prediction (IMP; see ref. 30) network analysis of DAF-16 neuronal target genes with STAM phenotypes (red circles).

*daf-2(e1370)* (Fig. 2d, e), both in whole-life and adult-only RNA interference (RNAi) tests. (Neuronal RNAi is effective in learning, STAM, and LTAM tests<sup>21</sup>.) The variety of genes (ion channels, transcription factors, G-proteins, vesicle fusion proteins) required for *daf-2* mutants' extended STAM suggests that decreased insulin signalling affects a broad network of memory extension genes. Several of these genes are also required for learning and memory in wild type (Extended Data Fig. 6g), suggesting that *daf-2* mutants maintain neuronal function, rather than using an alternative short-term memory mechanism.

*daf-2* mutants also maintain motor neuron axon regeneration ability with age in a *daf-16*-dependent manner<sup>3</sup>, and we found this is also true for mechanosensory neurons (Fig. 3a, b, Extended Data Fig. 7a–d). To identify factors that enable axon regeneration with age, we isolated and RNA-sequenced six adult mechanosensory neurons (Fig. 3c, Supplementary Table 9); this set includes 94 known larval regeneration genes from limited candidate screens<sup>22</sup> ( $P \leq 1.82 \times 10^{-20}$ ). To find *daf-2/daf-16*-dependent axon regeneration candidates, we identified mechanosensory genes that are also regulated by neuronal IIS/FOXO (Fig. 3c, Supplementary Table 9;  $P < 0.002$ ). The forkhead transcription factor FKH-9 is a neuronal IIS/FOXO target (Supplementary Table 6) and a canonical Class I target<sup>1</sup>, and is expressed in mechanosensory neurons (Supplementary Table 9). The *fkh-9* promoter is occupied by DAF-16, which we confirmed by chromatin immunoprecipitation followed by quantitative PCR (ChIP-qPCR; Fig. 3e, Extended Data Fig. 8a, b). FKH-9–GFP localized to nuclei,

and neurons were the primary site of differential FKH-9–GFP levels in *daf-2* mutants (Fig. 3f, Extended Data Fig. 8c), all suggesting a role for FKH-9 in *daf-2/daf-16*-mediated neuronal function.

While there is no effect on the first day of adulthood (Extended Data Fig. 7e, f), loss of *fkh-9* severely impairs axon regeneration ability in aged (day 5) *daf-2* mutants (Fig. 4a), correlating with an increased difference in *fkh-9* expression levels between wild-type and *daf-2* (Fig. 3d). Pan-neuronal *fkh-9* expression rescues the ability of day 5 *daf-2;fkh-9* worms to regenerate PLM axons (Fig. 4b, c). *fkh-9* levels are critical for neuron morphology, as *fkh-9* neuronal overexpression causes axonal defects (Extended Data Fig. 7g).

Adult-specific and whole-life reduction of *fkh-9* also severely impaired extended STAM of *daf-2* mutants (Fig. 4d, Extended Data Fig. 9). *daf-2;fkh-9* double mutants were defective in both STAM and learning, and neuronal *fkh-9* expression rescued these defects (Fig. 4e, Extended Data Fig. 9d, e), suggesting that *fkh-9* is required for extended memory and normal neuronal development in *daf-2* mutants. Day 1 and 5 *fkh-9* expression levels correlated with STAM and axon regeneration (Fig. 3d). *fkh-9* reduction delayed development, and reduction during adulthood caused severe matricide (Extended Data Fig. 10a–c). *fkh-9* knockdown in adult *daf-2* worms treated with FUdR (5-fluoro-2'-deoxyuridine) to block matricide<sup>20</sup> significantly shortened lifespan (40–50%; Fig. 4f). Pan-neuronal *fkh-9* expression did not rescue lifespan (Extended Data Fig. 10d), suggesting that FKH-9 acts in non-neuronal tissues to regulate lifespan.

Thus, IIS/FOXO-regulated FKH-9 function is important for both neuronal and non-neuronal growth and development, as well as adult memory and axon regeneration. Interestingly, the FKH-9 mammalian homologue FOXG1 is required for axon outgrowth<sup>23</sup> and is the most highly-induced gene in spinal cords treated with radial glial cell transplant following spinal cord injury<sup>24</sup>.

Network analysis using *fkh-9* and the other 8 neuronal DAF-16 STAM genes (Fig. 4g, Supplementary Table 10) identified *casy-1*, which is required for several forms of associative learning and memory<sup>2,25–27</sup>, *apl-1*, the *C. elegans* orthologue of amyloid precursor protein (APP) that can disrupt sensory plasticity<sup>28</sup>, and *dlk-1*, the only previously known regulator of age-dependent axon regeneration<sup>3,29</sup>. Additionally, genes involved in neuronal degeneration (*mec-17*), neuronal development (*egl-44*, *sem-4*), neuronal function (*egl-21*, *rcn-1*, *vab-9*, *cysl-1*), synaptic regulation and function (*cab-1*, *hlb-1*, *magu-4*, *sph-1*, *unc-64*), and axon outgrowth (*unc-14*) and regeneration (*egl-8*, *fos-1*, *pmk-3*), were connected to the STAM genes. PQM-1 (ref. 8), whose motif (DAE) is overrepresented in neuronal IIS target promoters, and other IIS (*akt-2*, *dct-6*, *hlh-30*), TGF- $\beta$  (*daf-14*, *sma-4*, *crm-1*, *sma-9*, *sma-1*, *sta-1*), and MAPK pathway (*vhp-1*, *pmk-3*) components emerged in the network. Transcriptional regulation by IIS/FOXO and its targets may lead to broader, indirect transcriptional and non-transcriptional regulation of genes with important neuronal functions.

Plasticity in development, reproduction and longevity allows organisms to respond appropriately to nutrient availability and changes in their environment. The IIS pathway is a critical mediator of these decisions, with FOXO selecting transcriptional targets to execute specific biochemical functions in each tissue, including factors that maintain cognitive function with age. *daf-2* mutant worms maintain neuronal behaviours with age by using a set of transcriptional targets that are distinct from previously identified metabolic and stress resistance targets expressed in other tissues. These genes may regulate additional neuronal targets through non-transcriptional mechanisms (Fig. 4g). The regulation of tissue-specific transcriptional programs is important to coordinate phenotypic responses, extending neuronal abilities in concert with the extended longevity and reproductive span of *daf-2* mutants.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 28 May; accepted 25 November 2015.**

**Published online 14 December 2015.**

- Murphy, C. T. *et al.* Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. *Nature* **424**, 277–283 (2003).
- Kauffman, A. L., Ashraf, J. M., Corces-Zimmerman, M. R., Landis, J. N. & Murphy, C. T. Insulin signaling and dietary restriction differentially influence the decline of learning and memory with age. *PLoS Biol.* **8**, e1000372 (2010).
- Byrne, A. B. *et al.* Insulin/IGF1 signaling inhibits age-dependent axon regeneration. *Neuron* **81**, 561–573 (2014).
- Sjöberg, J. & Kanje, M. Insulin-like growth factor (IGF-1) as a stimulator of regeneration in the freeze-injured rat sciatic nerve. *Brain Res.* **485**, 102–108 (1989).
- Wolkow, C. A., Kimura, K. D., Lee, M. S. & Ruvkun, G. Regulation of *C. elegans* life-span by insulinlike signaling in the nervous system. *Science* **290**, 147–150 (2000).
- Libina, N., Berman, J. R. & Kenyon, C. Tissue-specific activities of *C. elegans* DAF-16 in the regulation of lifespan. *Cell* **115**, 489–502 (2003).
- Luo, S., Kleemann, G. A., Ashraf, J. M., Shaw, W. M. & Murphy, C. T. TGF- $\beta$  and insulin signaling regulate reproductive aging via oocyte and germline quality maintenance. *Cell* **143**, 299–312 (2010).
- Tepper, R. G. *et al.* PQM-1 complements DAF-16 as a key transcriptional regulator of DAF-2-mediated development and longevity. *Cell* **154**, 676–690 (2013).
- Zhang, P., Judy, M., Lee, S.-J. & Kenyon, C. Direct and indirect gene regulation by a life-extending FOXO protein in *C. elegans*: roles for GATA factors and lipid gene regulators. *Cell Metab.* **17**, 85–100 (2013).
- Murakami, H., Bessinger, K., Hellmann, J. & Murakami, S. Aging-dependent and -independent modulation of associative learning behavior by insulin/insulin-like growth factor-1 signal in *Caenorhabditis elegans*. *J. Neurosci.* **25**, 10894–10904 (2005).

- Tank, E. M. H., Rodgers, K. E. & Kenyon, C. Spontaneous age-related neurite branching in *Caenorhabditis elegans*. *J. Neurosci.* **31**, 9279–9288 (2011).
- Toth, M. L. *et al.* Neurite sprouting and synapse deterioration in the aging *Caenorhabditis elegans* nervous system. *J. Neurosci.* **32**, 8778–8790 (2012).
- Chikina, M. D., Huttenhower, C., Murphy, C. T. & Troyanskaya, O. G. Global prediction of tissue-specific gene expression and context-dependent gene networks in *Caenorhabditis elegans*. *PLOS Comput. Biol.* **5**, e1000417 (2009).
- Ewald, C. Y., Landis, J. N., Abate, J. P., Murphy, C. T. & Blackwell, T. K. Dauer-independent insulin/IGF-1-signalling implicates collagen remodelling in longevity. *Nature* **519**, 97–101 (2015).
- Zhang, S., Banerjee, D. & Kuhn, J. R. Isolation and culture of larval cells from *C. elegans*. *PLoS ONE* **6**, e19505 (2011).
- Spencer, W. C. *et al.* Isolation of specific neurons from *C. elegans* larvae for gene expression profiling. *PLoS ONE* **9**, e112102 (2014).
- Schafer, J. C. *et al.* IFTA-2 is a conserved cilia protein involved in pathways regulating longevity and dauer formation in *Caenorhabditis elegans*. *J. Cell Sci.* **119**, 4088–4100 (2006).
- Apfeld, J. & Kenyon, C. Regulation of lifespan by sensory perception in *Caenorhabditis elegans*. *Nature* **402**, 804–809 (1999).
- Hansen, M., Hsu, A.-L., Dillin, A. & Kenyon, C. New genes tied to endocrine, metabolic, and dietary regulation of lifespan from a *Caenorhabditis elegans* genomic RNAi screen. *PLoS Genet.* **1**, 119–128 (2005).
- Shaw, W. M., Luo, S., Landis, J., Ashraf, J. & Murphy, C. T. The *C. elegans* TGF- $\beta$  Dauer pathway regulates longevity via insulin signaling. *Curr. Biol.* **17**, 1635–1645 (2007).
- Lakhina, V. *et al.* Genome-wide functional analysis of CREB/long-term memory-dependent transcription reveals distinct basal and memory gene expression programs. *Neuron* **85**, 330–345 (2015).
- Chen, L. *et al.* Axon regeneration pathways identified by systematic genetic screening in *C. elegans*. *Neuron* **71**, 1043–1057 (2011).
- Tian, N. M., Pratt, T. & Price, D. J. Foxg1 regulates retinal axon pathfinding by repressing an ipsilateral program in nasal retina and by causing optic chiasm cells to exert a net axonal growth-promoting activity. *Development* **135**, 4081–4089 (2008).
- Chang, Y.-W. *et al.* Rapid induction of genes associated with tissue protection and neural development in contused adult spinal cord after radial glial cell transplantation. *J. Neurotrauma* **26**, 979–993 (2009).
- Ikeda, D. D. *et al.* CASY-1, an ortholog of calyntenin/alcaideins, is essential for learning in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **105**, 5260–5265 (2008).
- Hoerndli, F. J. *et al.* A conserved function of *C. elegans* CASY-1 calyntenin in associative learning. *PLoS ONE* **4**, e4880 (2009).
- Ohno, H. *et al.* Role of synaptic phosphatidylinositol 3-kinase in a behavioral learning response in *C. elegans*. *Science* **345**, 313–317 (2014).
- Ewald, C. Y. *et al.* Pan-neuronal expression of APL-1, an APP-related protein, disrupts olfactory, gustatory, and touch plasticity in *Caenorhabditis elegans*. *J. Neurosci.* **32**, 10156–10169 (2012).
- Hammamund, M., Nix, P., Hauth, L., Jorgensen, E. M. & Bastiani, M. Axon regeneration requires a conserved MAP kinase pathway. *Science* **323**, 802–806 (2009).
- Wong, A. K., Krishnan, A., Yao, V., Tadych, A. & Troyanskaya, O. G. IMP 2.0: a multi-species functional genomics portal for integration, visualization and prediction of protein functions and networks. *Nucleic Acids Res.* **43**, W128–W133 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank the *C. elegans* Genetics Center for strains; WormBase (WS250); L. Parsons for RNA-seq data support; J. Wiggins and the Lewis-Sigler High Throughput Sequencing Core Facility for RNA-seq support; C. DeCoste and the Flow Cytometry Facility for assistance; V. Yao for tissue prediction analysis; R. DiLoreto for chemotaxis assay analysis; and Z. Gitai, the Murphy lab, and W. Mair for discussion. Funding was provided by a Keck Scholars Program fellowship (C.T.M.), National Institutes of Health Cognitive Aging R01 (C.T.M.), Ruth L. Kirschstein National Research Service Awards (R.K., R.A.), National Science Foundation (J.L.) and New Jersey Commission on Brain Injury Research (V.L.) fellowships.

**Author Contributions** C.T.M., R.K., V.L. and J.L. designed experiments. R.K., V.L., R.A., J.L., J.A., and C.T.M. performed experiments and analysed data. R.K., V.L. and J.L. performed tissue isolation and RNA-seq. A.W., R.K., V.L. and J.L. performed bioinformatics analysis. R.K., V.L. and R.A. performed short-term memory experiments. V.L. performed axon regeneration experiments. R.K., V.L. and C.T.M. wrote the manuscript. R.A. and A.W. contributed equally to this work.

**Author Information** Sequencing reads are deposited at NCBI BioProject under accession number PRJNA297798. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.T.M. ([ctmurphy@princeton.edu](mailto:ctmurphy@princeton.edu)).

## METHODS

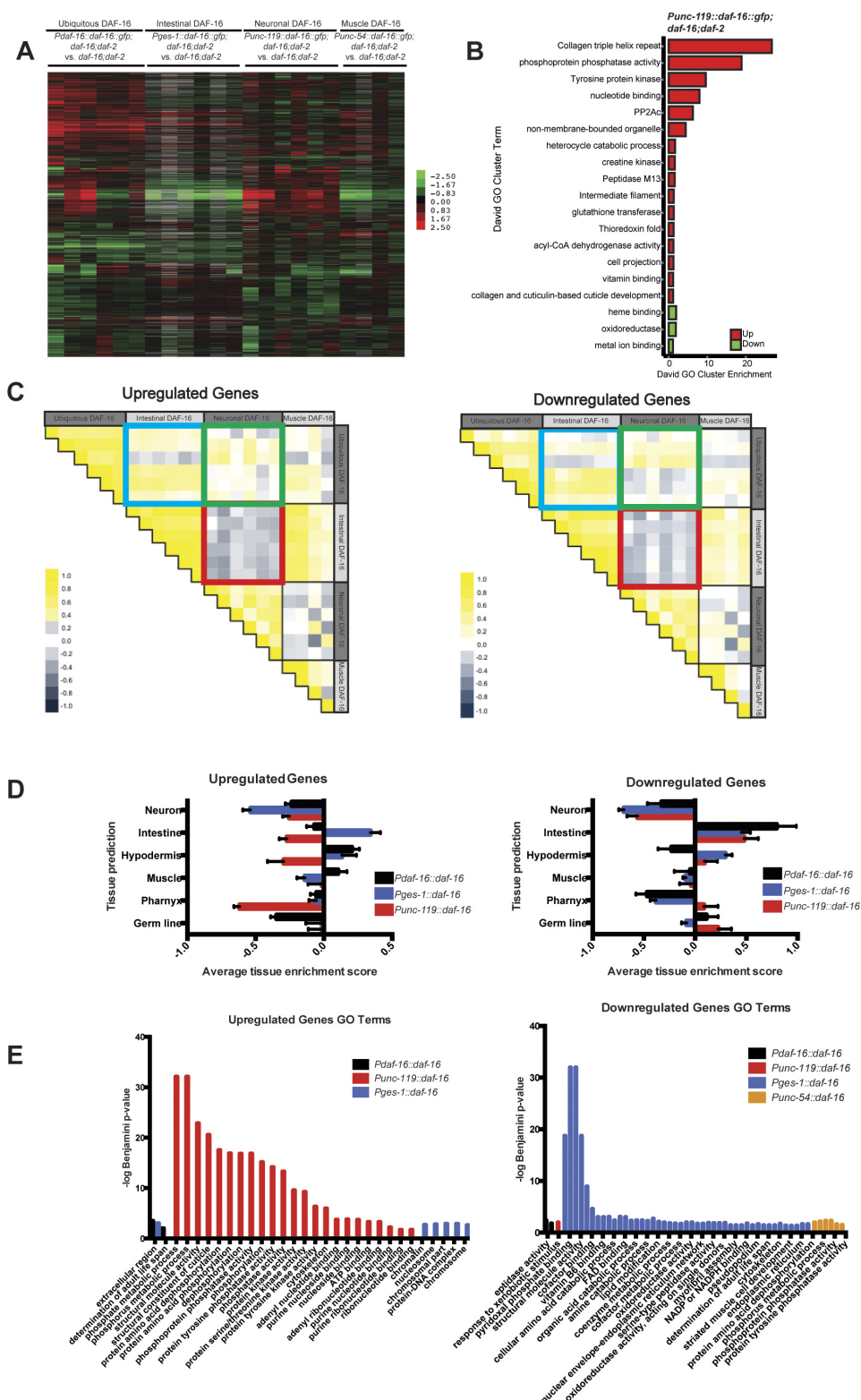
**Adult cell isolation.** Day 1 adult neuronally GFP-labelled worms (*Punc119::GFP* or *Pmec-4::GFP*) were prepared for cell isolation as previously described<sup>15</sup> with modifications (Extended Data Fig. 2). Synchronized adult worms were washed with M9 buffer to remove excess bacteria. The pellet (~250 µl) was washed with 500 µl lysis buffer (200 mM DTT, 0.25% SDS, 20 mM HEPES pH 8.0, 3% sucrose) and resuspended in 1,000 µl lysis buffer. Worms were incubated in lysis buffer with gentle rocking for 6.5 min at room temperature. The pellet was washed 6× with M9 and resuspended in 20 mg ml<sup>-1</sup> pronase from *Streptomyces griseus* (Sigma-Aldrich). Worms were incubated at room temperature (<20 min) with periodic mechanical disruption by pipetting every 2 min. When most worm bodies were dissociated, leaving only small debris and eggs, ice-cold PBS buffer containing 2% fetal bovine serum (Gibco) was added. RNA from FACS-sorted neurons was prepared for RNA-seq and subsequent analysis (see Extended Data for details).

No statistical methods were used to predetermine sample size. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

**Short-term associative memory assay.** Memory assays were performed as described<sup>2</sup>.

**Axon regeneration assays.** *In vivo* laser axotomy of PLM neurons was performed as described<sup>22</sup>.

31. Tomioka, M. *et al.* The insulin/PI 3-kinase pathway regulates salt chemotaxis learning in *Caenorhabditis elegans*. *Neuron* **51**, 613–625 (2006).
32. Kodama, E. *et al.* Insulin-like signaling and the neural circuit for integrative behavior in *C. elegans*. *Genes Dev.* **20**, 2955–2960 (2006).
33. Murakami, H., Bessinger, K., Hellmann, J. & Murakami, S. Aging-dependent and -independent modulation of associative learning behavior by insulin/insulin-like growth factor-1 signal in *Caenorhabditis elegans*. *J. Neurosci.* **25**, 10894–10904 (2005).
34. Stein, G. M. & Murphy, C. T. C. *C. elegans* positive olfactory associative memory is a molecularly conserved behavioral paradigm. *Neurobiol. Learn. Mem.* **115**, 86–94 (2014).

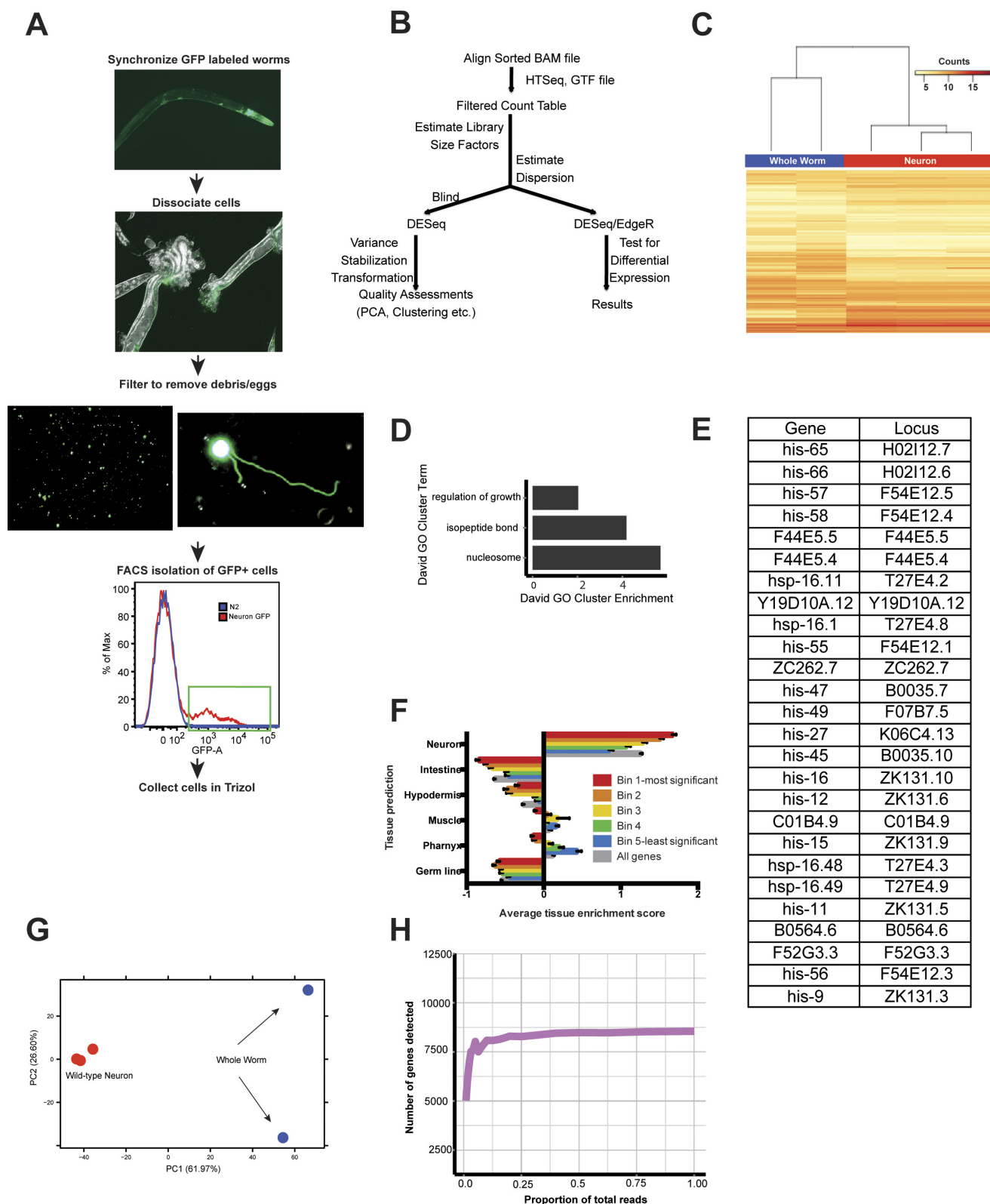


### Extended Data Figure 1 | Tissue-specific rescue of DAF-16 activity in *daf-16;daf-2* mutants identifies distinct gene expression profiles.

**a**, DAF-16 tissue-specific transgenics; heatmap of all genes with expression differences  $\geq 1.5$ -fold in  $\geq 3$  arrays. **b**, Significant gene ontology (GO) cluster terms from *Punc-119::daf-16*-regulated up- and downregulated genes (enrichment score  $> 1$ ). **c**, Pairwise Pearson correlations between arrays of DAF-16-upregulated or downregulated targets. The red box highlights the negative correlation between neuronal DAF-16 rescued targets (*Punc-119::daf-16:gfp;daf-16;daf-2* vs *daf-16;daf-2*) and intestinal DAF-16 targets (*Pges-1::daf-16:gfp;daf-16;daf-2* vs *daf-16;daf-2*), while the blue box shows the positive correlation between intestinal DAF-16 targets

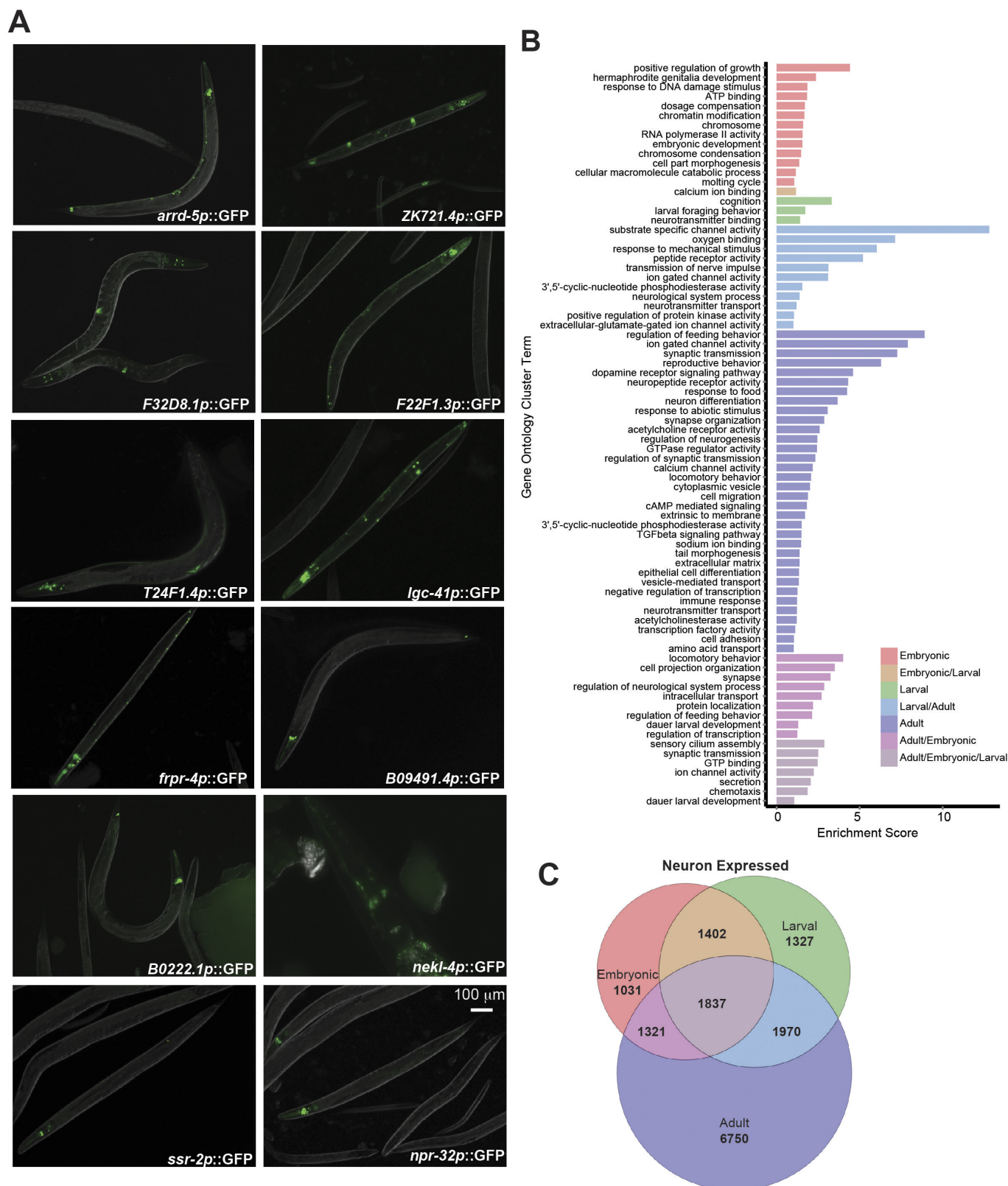
(*Pges-1::daf-16:gfp;daf-16;daf-2* vs *daf-16;daf-2*) and whole-worm DAF-16 targets (*Pdaf-16::daf-16:gfp;daf-16;daf-2* vs *daf-16;daf-2*). The green box shows the weak correlation between neuronal rescued and whole-worm DAF-16 targets. **d**, Tissue enrichment analysis (mean  $\pm$  s.e.m.) of significant DAF-16-rescued up- and downregulated genes (Supplementary Table 1) (FDR  $< 0.5$ ). **e**, Significant GO terms (adjusted *P* value  $< 0.05$ ) for DAF-16 upregulated and downregulated genes from whole worm, intestine-, neuron- and muscle-rescued DAF-16 strains. Genes used for GO analysis (Supplementary Table 2) were derived from SAM analysis of the microarrays in **a** and Supplementary Table 1.





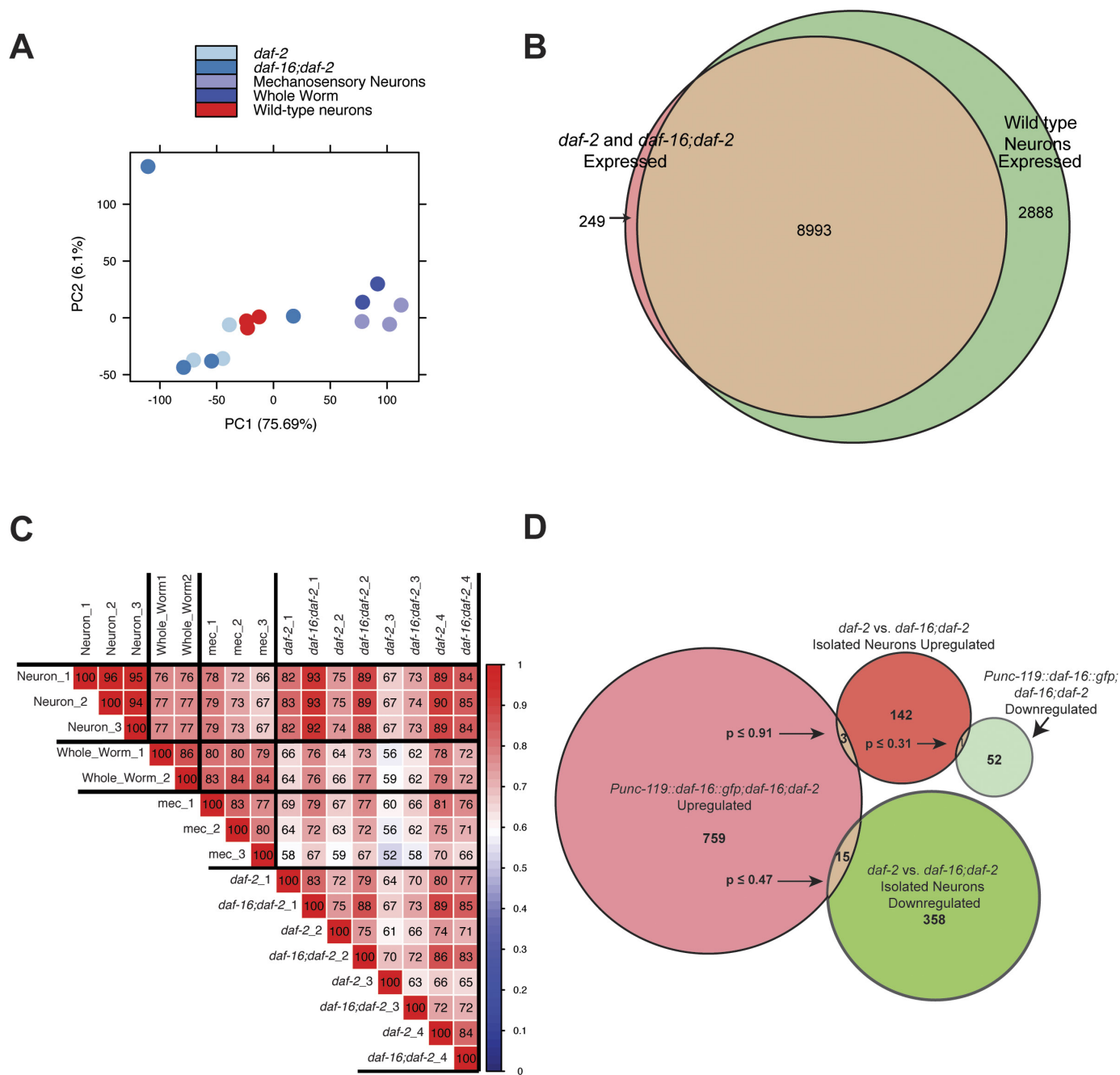
**Extended Data Figure 2 | Protocol for isolating neuron-specific targets using FACS followed by RNA-sequencing.** **a**, Pipeline for isolation of adult cells for FACS and RNA sequencing. **b**, Workflow for RNA-seq data analysis of isolated neurons. **c**, Heatmap of wild-type neuron-expressed relative to whole-worm-expressed genes. **d**, Actinomycin D (transcription inhibitor) treatment ( $100\mu\text{g ml}^{-1}$ ) during the cell isolation process demonstrates that the neuron isolation technique induces minimal transcriptional changes in wild type animals. Gene ontology (GO) terms represent genes upregulated in the absence of actinomycin D (Fig. 1b, Supplementary Table 4). **e**, The 26 differentially expressed genes from

actinomycin D treatment are listed. **f**, *C. elegans* tissue gene expression prediction confirms neuronal character of adult wild-type neuron-enriched genes. Neuron-enriched genes were divided among equal bins according to *P* value. Bin 1: FDR < 0.003%; bin 2: 0.003–0.03%; bin 3: 0.03–1.3%; bin 4: 1.3–4%; bin 5: 4–10%. **g**, Principal component analysis (PCA) shows a clear separation between wild-type adult neuronal and whole-worm samples. **h**, Downsampling of wild-type neuron sequencing reads demonstrates sufficient sampling depth. The number of genes detected at the 3 counts per million threshold (for expressed genes) with different proportions of total sequencing depth analysed.



**Extended Data Figure 3 | Neuron-expressed genes identified by our method are confirmed to be expressed in adults and have adult neuronal functions.** **a**, Promoter–GFP transcriptional fusions of candidate uncharacterized neuronal genes (day 1 of adulthood). **b**, Gene ontology clusters were generated from the categories in Fig. 1e. Non-overlapping GO terms suggest a transition from development-related processes in

embryonic and larval animals to neuronal processes involved in behaviour in adults (Supplementary Table 5). **c**, Venn diagram depicting the overlap between genes classified as “expressed” among embryonic and larval neurons<sup>16</sup> and adult neurons from our RNA-seq analysis (Supplementary Table 5).



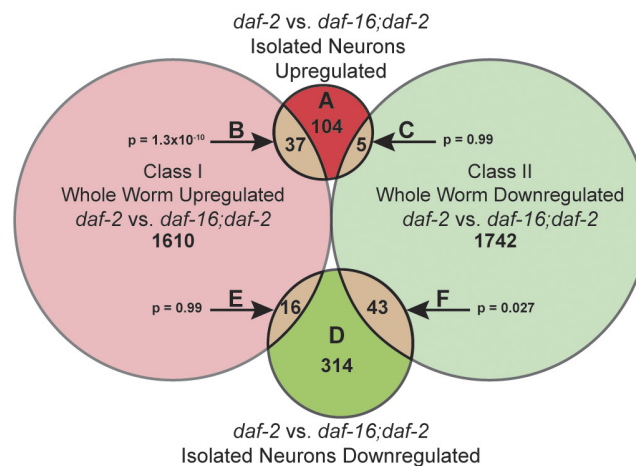
**Extended Data Figure 4 | Comparison of neuronal DAF-16 targets with wild-type neuronal targets and whole-worm DAF-16 targets.** **a**, Principal component analysis of the whole worm and isolated adult neuron samples obtained for this study. **b**, Venn diagram depicting the overlap of *daf-2*- and *daf-16;daf-2*-expressed genes with those expressed in wild-type adult neurons. **c**, Spearman correlation of whole-worm and isolated adult neuron samples. **d**, The DAF-16 cell-autonomous

and cell-non-autonomous targets are distinct. The number of genes that overlap between neuronal DAF-16-rescued whole-worm targets (*Punc-119::daf-16::gfp;daf-16;daf-2* vs *daf-16;daf-2*) and isolated neuron IIS targets (*daf-2* vs *daf-16;daf-2*) is shown (Supplementary Table 8). Hypergeometric distribution analysis (*P* values) shows that the extent of overlap between the gene categories is not significant.

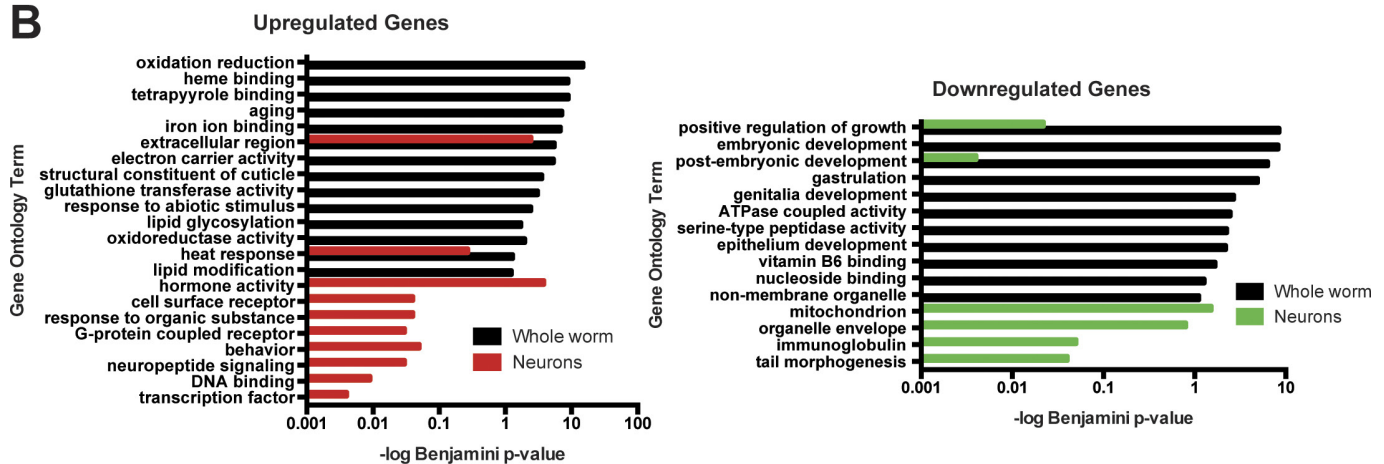
A

Group	Neuron IIS	Description	% DBE	% DAE
A	Neuron IIS Upregulated	No Overlap with Whole Worm IIS	22.1	21.2
B		Overlap with whole worm IIS upregulated	43.2	21.6
C		Overlap with IIS whole worm downregulated	20	0
D	Neuron IIS Downregulated	No Overlap with Whole Worm IIS	28.2	15.1
E		Overlap with whole worm IIS upregulated	25	18.8
F		Overlap with whole worm IIS downregulated	27.9	27.9

Whole genome occurrence: DBE 30.7%, DAE 20.8%



B

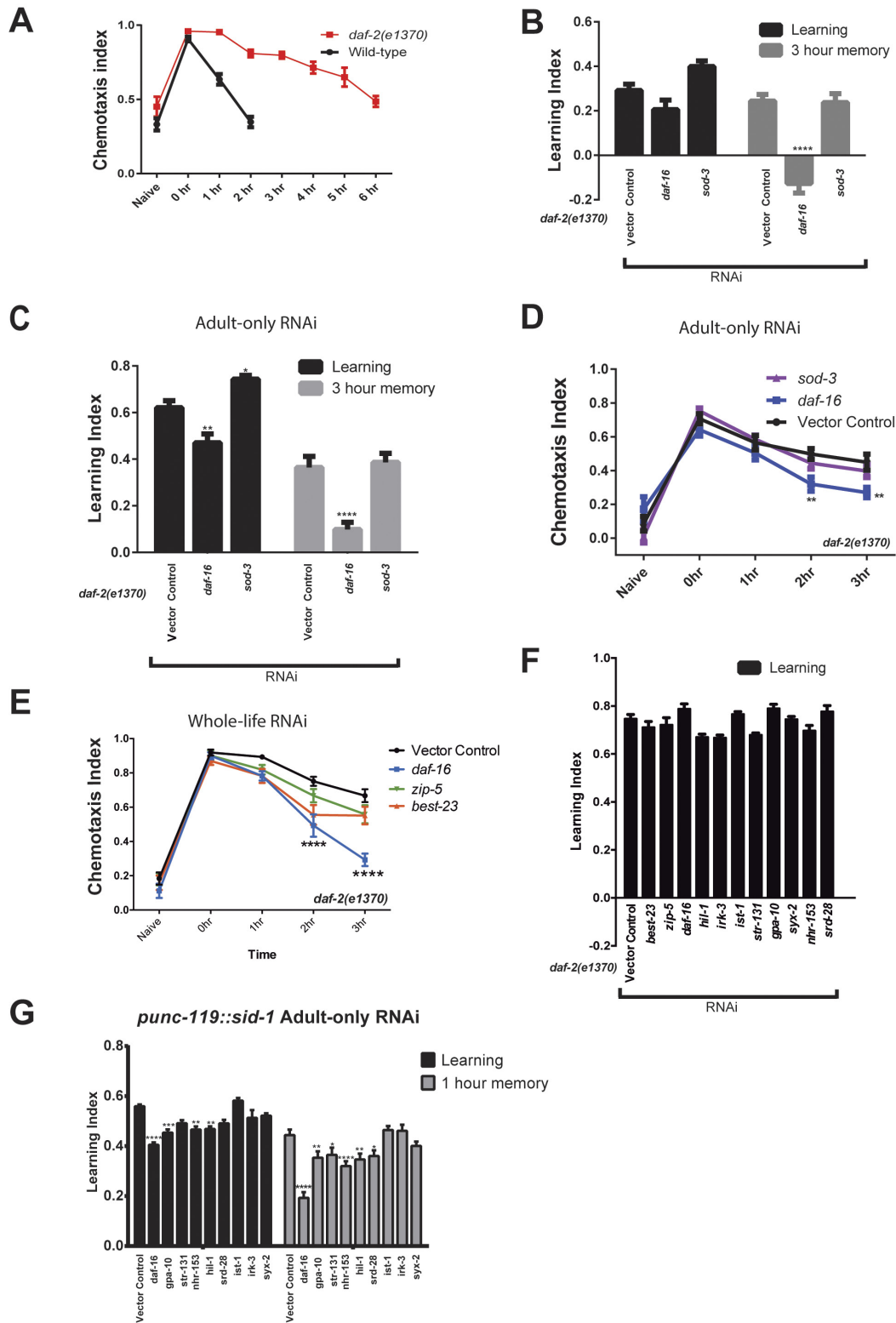


**Extended Data Figure 5 | Promoter analysis and gene ontology term analysis of neuronal IIS/FOXO genes.** a, The different classes of neuronal IIS/FOXO genes shown in Fig. 2b were analysed for DBE and DAE sequences in the 1 kb upstream promoter regions. The genome-wide percentage of DBE and DAE occurrences across the 1 kb promoters of all

gene-encoding regions is reported. Comparison of whole-worm (Class I)<sup>8</sup> vs neuronal-IIS/FOXO-regulated targets. *P* values: hypergeometric distributions. b, GO terms of Class I whole worm<sup>8</sup> vs neuronal-IIS upregulated genes (left) and Class II whole worm<sup>8</sup> vs neuronal-IIS downregulated genes (right) (Supplementary Table 5).



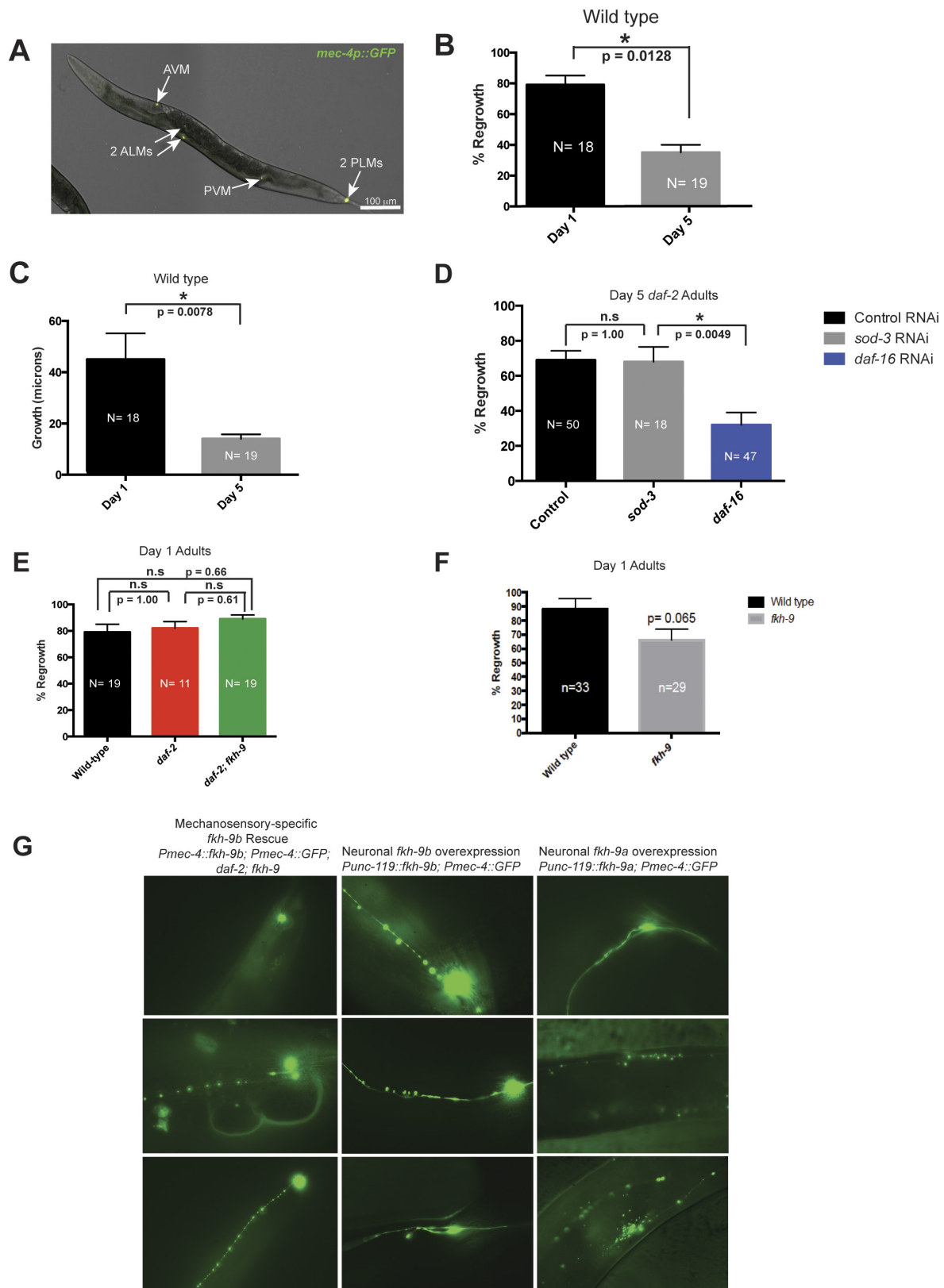
## Whole-life RNAi



Extended Data Figure 6 | See next page for caption.

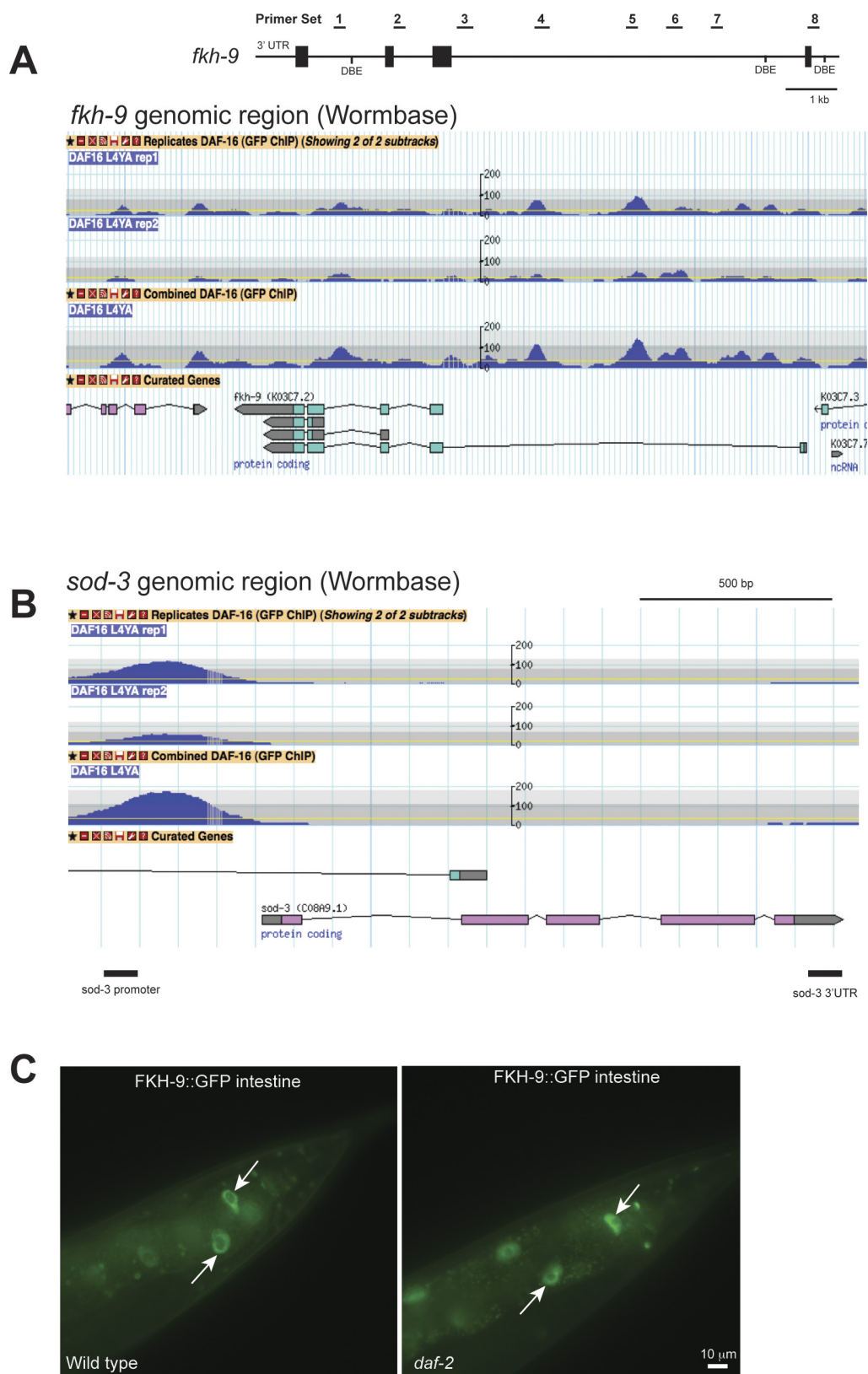
**Extended Data Figure 6 | Short-term associative memory phenotypes obtained upon knocking down neuronal IIS genes in *daf-2* mutants and wild-type animals.** *daf-2* is required for various forms of *C. elegans* associative learning<sup>2,27,31–34</sup>. *daf-16* is required for the improvements and extensions of abilities with age of *daf-2* mutants<sup>2</sup>. *daf-2* mutants are defective for salt chemotaxis learning<sup>27,31,32</sup>, and *daf-16* is not involved in salt chemotaxis learning<sup>27,31,32</sup>. Furthermore, salt learning utilizes a unique *daf-2c* isoform<sup>27</sup> in a *daf-16*-independent manner<sup>31</sup>, suggesting a learning mechanism distinct from the associative memory paradigms studied here. We are specifically interested in understanding how activation of DAF-16 results in the improved and extended abilities of *daf-2* mutants to carry out olfactory associative learning<sup>2</sup>, short-term associative memory<sup>2,34</sup>, and long-term associative memory<sup>2</sup>, all of which require *daf-16*. **a**, Chemotaxis index profile of wild type (N2) and *daf-2* animals at time points following memory training. **b**, RNAi knockdown of *sod-3*, a non-neuronal DAF-16-regulated target that influences lifespan, has no effect on the extended

short-term associative memory (STAM) of *daf-2* mutants when treated with RNAi-feeding bacteria throughout the whole life (**b**) or only the post-developmental (adult-only) period (**c**, **d**) of the animal. *daf-2* worms treated with *daf-16* RNAi have defective STAM, as previously reported<sup>2</sup>. **e**, Knockdown of the neuronal IIS candidate genes *zip-5* and *best-23* does not affect STAM. Time-courses showing the chemotaxis index for each time point are shown in **d** and **e**. Learning indices are shown in **b**, **c**, **f** and **g**. **b–e**, Two-way repeated measures ANOVA, Bonferroni post hoc tests. **f**, Treatment of *daf-2* worms with neuronal DAF-16 target RNAi does not affect short-term associative learning. **g**, Neuronal-RNAi sensitive worms (*Punc-119::sid-1*) in a wild-type background were treated only during adulthood with RNAi targeted against the neuronal DAF-16 target genes. Learning (0 h) and 1 h short-term associative memory time points are shown. **a–g**, Mean  $\pm$  s.e.m., \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$ .



**Extended Data Figure 7 | Characterization of age-dependent axon regeneration and structural defects upon *fkh-9* overexpression in mechanosensory neurons.** **a**, Six adult mechanosensory neurons labelled by *mec-4p::GFP* were isolated for RNA-seq. **b**, Axon length from the cell body to the site of injury was measured in  $\mu\text{m}$  immediately after axotomy and 24 h later. Regenerative capacity of wild-type PLM axons declines from day 1 to day 5 of adulthood. **c**, Day 5 wild-type animals regrow axons that are significantly shorter than in day 1 animals. **d**, Axotomies of *daf-2* mutants grown on vector control, *sod-3*, or *daf-16* RNAi demonstrate that

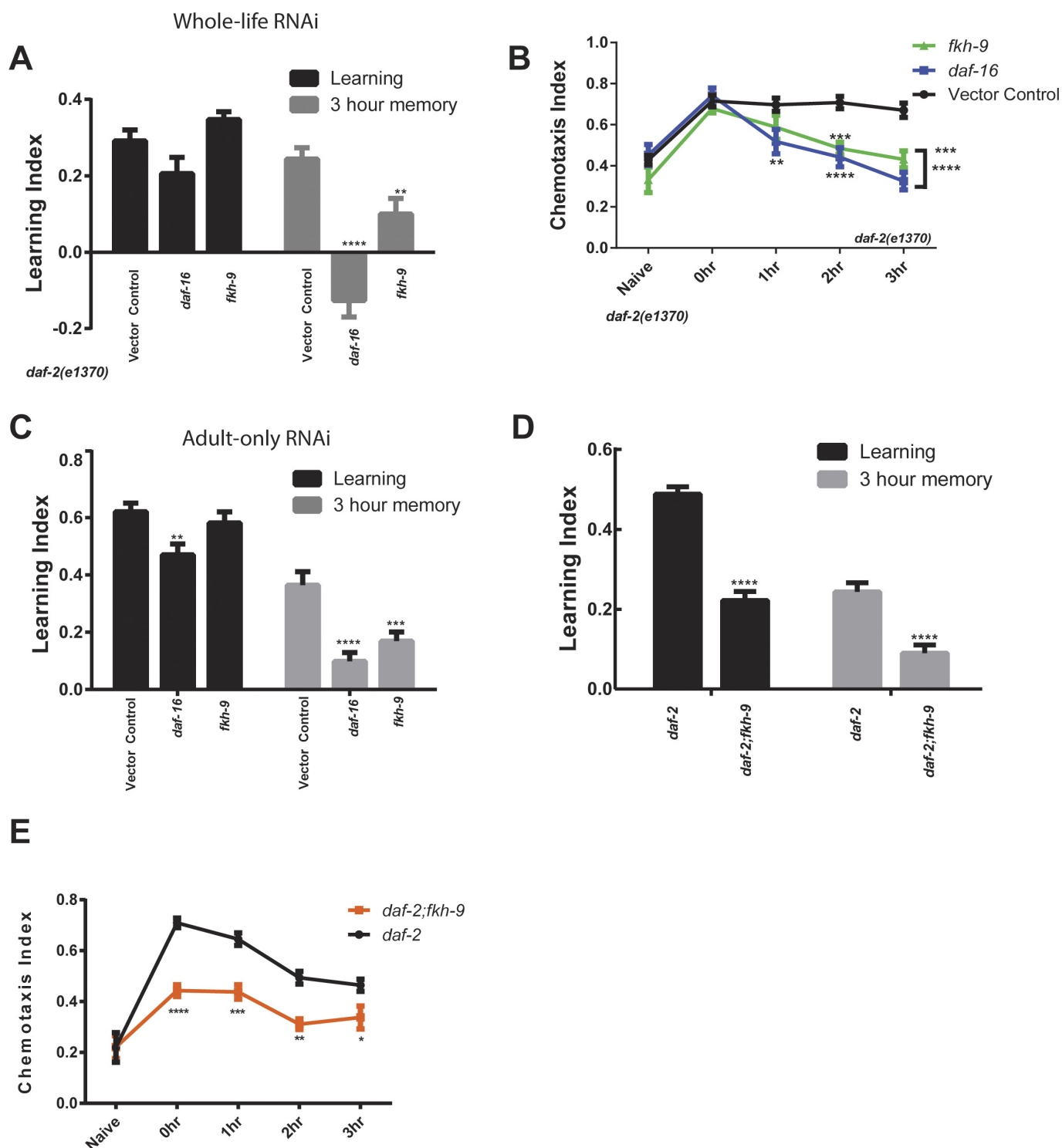
*sod-3*, a lifespan-regulating DAF-16 target, does not influence the axon regeneration capacity of *daf-2* worms at day 5 of adulthood. **e**, *fkh-9* does not affect the regenerative capacity of *daf-2* axons on day 1 of adulthood. **f**, *fkh-9* is not required for axon regeneration in day 1 adults. **b–f**, Mean  $\pm$  s.e.m., Fisher's exact test,  $*P < 0.05$ . **g**, Overexpression of the a and b isoforms of *fkh-9* in wild-type animals causes axonal structural defects. Rescuing *fkh-9* activity in the mechanosensory neurons of *daf-2;fkh-9* mutants results in severe beading and degeneration of axons.



**Extended Data Figure 8 |** WormBase gene models for *fkh-9* and *sod-3* are shown with modENCODE data for DAF-16 ChIP-seq experiments. **a, b,** Wormbase (<http://www.wormbase.org>) gene models for *fkh-9* (**a**) and *sod-3* (**b**). Primer sets for ChIP-qPCR are depicted in **a**. **c,** Posterior

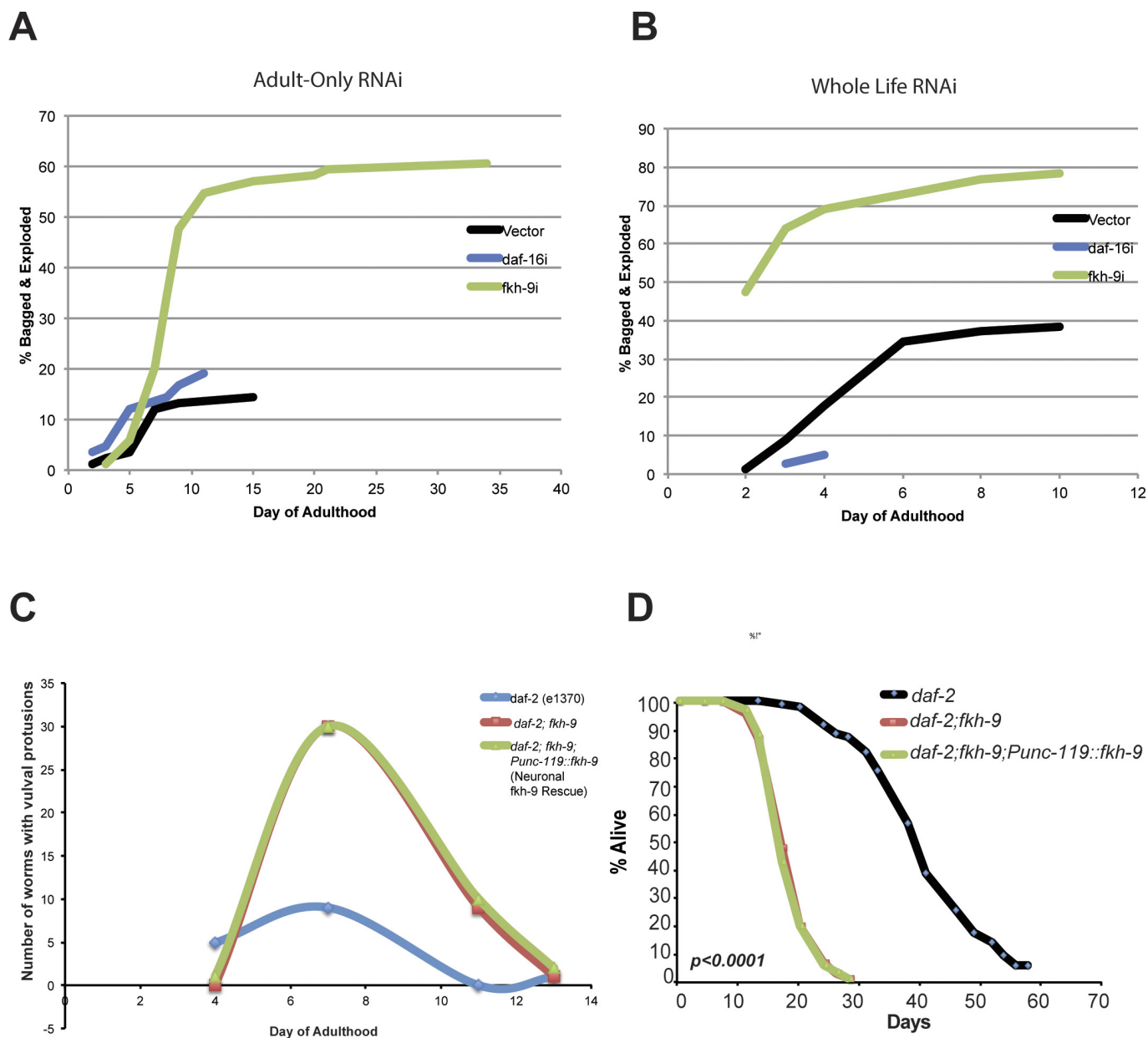
intestinal FKH-9-GFP expression is only modestly increased in *daf-2* compared to wild-type animals expressing *fkh-9p::fkh-9::gfp*.  $N = 25$  animals.





**Extended Data Figure 9 | Knocking down *fkh-9* via RNAi or using mutants reduces the enhanced short-term memory of *daf-2* animals.** **a, b**, Whole-life RNAi of *fkh-9* reduces *daf-2* STAM. **c**, RNAi knockdown of *fkh-9* exclusively during adulthood results in reduced *daf-2* STAM comparable to *daf-16* RNAi-treatment. **d, e**, *daf-2;fkh-9* mutants have

reduced learning (tested immediately following STAM training) and STAM compared to *daf-2*. Mean  $\pm$  s.e.m., \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$ . Time-courses showing the chemotaxis index for each time point are shown in **b** and **e**. Learning indices are shown in **a**, **c** and **d**.



**Extended Data Figure 10 | Neuronal FKX-9 is not required for the enhanced lifespan of *daf-2* mutants.** **a, b**, Adult-only (**a**) or whole-life (**b**) *fkh-9* RNAi treatment increases matricide in *daf-2* worms. The cumulative percentage of animals dead as a result of bagging and/or exploding was recorded every other day. Two biological replicates were performed, with a representative experiment shown. **c**, Neuronal rescue of *fkh-9* in *daf-2;fkh-9* animals does not diminish the rate of vulval protrusions with

age.  $N \geq 60$  per conditions for each experiment. **d**, Neuronal rescue of *fkh-9* does not restore longevity of the *daf-2;fkh-9* double mutant. *daf-2* median lifespan: 41 days, *daf-2;fkh-9* 20 days, *daf-2;fkh-9;Punc-119::fkh-9* 20 days.  $P < 0.0001$  for *daf-2* vs both *daf-2;fkh-9* and *daf-2;fkh-9;Punc-119::fkh-9*.  $N = 112$  worms per strain. Censor rate for *daf-2* 19%, *daf-2;fkh-9* 51%, *daf-2;fkh-9;Punc-119::fkh-9* 56%.

# Targeting *PTPRK*–*RSPO3* colon tumours promotes differentiation and loss of stem-cell function

Elaine E. Storm<sup>1</sup>, Steffen Durinck<sup>2</sup>, Felipe de Sousa e Melo<sup>1</sup>, Jarrod Tremayne<sup>3</sup>, Noelyn Kljavin<sup>1</sup>, Christine Tan<sup>4</sup>, Xiaofen Ye<sup>5</sup>, Cecilia Chiu<sup>4</sup>, Thinh Pham<sup>6</sup>, Jo-Anne Hongo<sup>4</sup>, Travis Bainbridge<sup>7</sup>, Ron Firestein<sup>6</sup>, Elizabeth Blackwood<sup>3</sup>, Ciara Metcalfe<sup>3</sup>, Eric W. Stawiski<sup>2</sup>, Robert L. Yauch<sup>5</sup>, Yan Wu<sup>4</sup> & Frederic J. de Sauvage<sup>1</sup>

**Colorectal cancer remains a major unmet medical need, prompting large-scale genomics efforts in the field to identify molecular drivers for which targeted therapies might be developed<sup>1–3</sup>. We previously reported the identification of recurrent translocations in R-spondin genes present in a subset of colorectal tumours<sup>4</sup>. Here we show that targeting *RSPO3* in *PTPRK*–*RSPO3*-fusion-positive human tumour xenografts inhibits tumour growth and promotes differentiation. Notably, genes expressed in the stem-cell compartment of the intestine were among those most sensitive to anti-*RSPO3* treatment. This observation, combined with functional assays, suggests that a stem-cell compartment drives *PTPRK*–*RSPO3* colorectal tumour growth and indicates that the therapeutic targeting of stem-cell properties within tumours may be a clinically relevant approach for the treatment of colorectal tumours.**

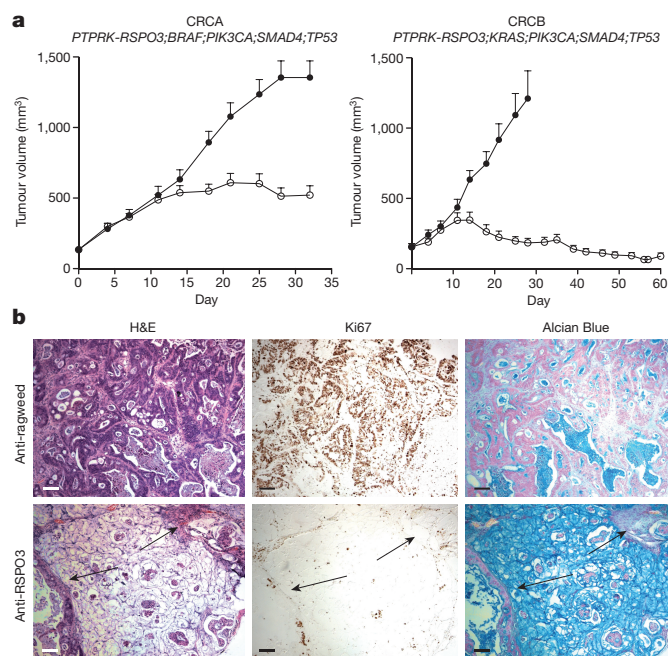
Molecular characterization of colorectal cancer (CRC) has revealed that the vast majority of tumours exhibit Wnt pathway activation that is largely driven through mutation of downstream signalling components<sup>1</sup>. Despite this knowledge, the development of therapies targeting the Wnt pathway has been challenging and CRC remains one of the top three most prevalent and deadly cancers<sup>5</sup>. We previously reported the identification of recurrent gene fusions in *RSPO2* and *RSPO3* in colorectal tumours<sup>4</sup>. R-spondin fusion tumours exhibited Wnt pathway activation, but did not contain common Wnt pathway mutations, such as *APC*<sup>4</sup>. R-spondins are known to amplify Wnt signalling<sup>6–8</sup>, suggesting that elevated expression resulting from translocation may drive Wnt-dependent tumour growth.

To test this hypothesis, we generated specific, function-blocking antibodies against *RSPO2* and *RSPO3* (Extended Data Fig. 1). We next screened patient-derived xenograft (PDX) samples and identified two *RSPO3*-fusion models, CRCA and CRCB, that harbour the *PTPRK*(*e1*)–*RSPO3*(*e2*) gene fusion<sup>4</sup>. These models do not contain common downstream mutations in the Wnt pathway, but do harbour *BRAF*(V600E), *PIK3CA*(E545K), *SMAD4*(D537E) and *TP53*(I195T) mutations (CRCA) and *KRAS*(G12V), *PIK3CA*(E545K), *SMAD4*(C499Y) and *TP53*(W53\*) mutations (CRCB). Both models express elevated *RSPO3*, but do not express *RSPO1*, 2 or 4 (Extended Data Fig. 2). This is similar to what is observed in *RSPO3*-fusion colon tumours, but is in contrast to the normal colon where both *RSPO2* and *RSPO3* are expressed (Extended Data Fig. 2).

Treatment with anti-*RSPO3* inhibited tumour growth in both models, demonstrating that the oncogenic driver of these tumours is *RSPO3* (Fig. 1a). Notably, tumours continued to grow for about one week, followed by either stasis (CRCA), or regression (CRCB), which persisted for at least 30 days (Extended Data Fig. 2). To characterize the static response of CRCA, we analysed end-of-study samples. Tumours from the control group continued to proliferate with evidence of secretory cell differentiation (Fig. 1b). By contrast, anti-*RSPO3*-treated tumours

exhibited large regions primarily composed of mucus. Epithelial cells that remained had a differentiated appearance and strongly reduced Ki67 positivity (Fig. 1b, arrows). The accumulation of mucus suggests measurements may be overestimating tumour content and that stasis may be the result of differentiation.

To gain insight into the mechanism through which anti-*RSPO3* promoted tumour growth inhibition, we first performed histopathological analysis 4 days after treatment initiation. Tumours from both models were similarly organized and highly proliferative, with ongoing differentiation, as revealed by KRT20 and MUC2 staining (Extended Data Fig. 3). There were some apparent differences between the models. MUC2 staining was largely extracellular in CRCA, whereas intracellular MUC2 was apparent in CRCB (Extended Data Fig. 3, arrows). In addition, Ki67



**Figure 1 | Anti-*RSPO3* inhibits tumour growth of *RSPO3*-fusion-positive PDX models. a**, Tumour growth inhibition in CRCA and CRCB following treatment with antibodies at 30 mg kg<sup>−1</sup>, twice a week, for three (CRCA) or four (CRCB) weeks. Closed circles, anti-ragweed treated. Open circles, anti-*RSPO3*-treated. Data represented as means ± s.e.m. *n* = 9 CRCB ragweed and *n* = 10 for all other groups. **b**, Sections stained as indicated from representative tumour samples of CRCA collected at the end of study. Arrows indicate tumour cells that appear differentiated. Scale bars, 100 µm. Xenograft experiments were performed at least two independent times. H&E, haematoxylin and eosin.

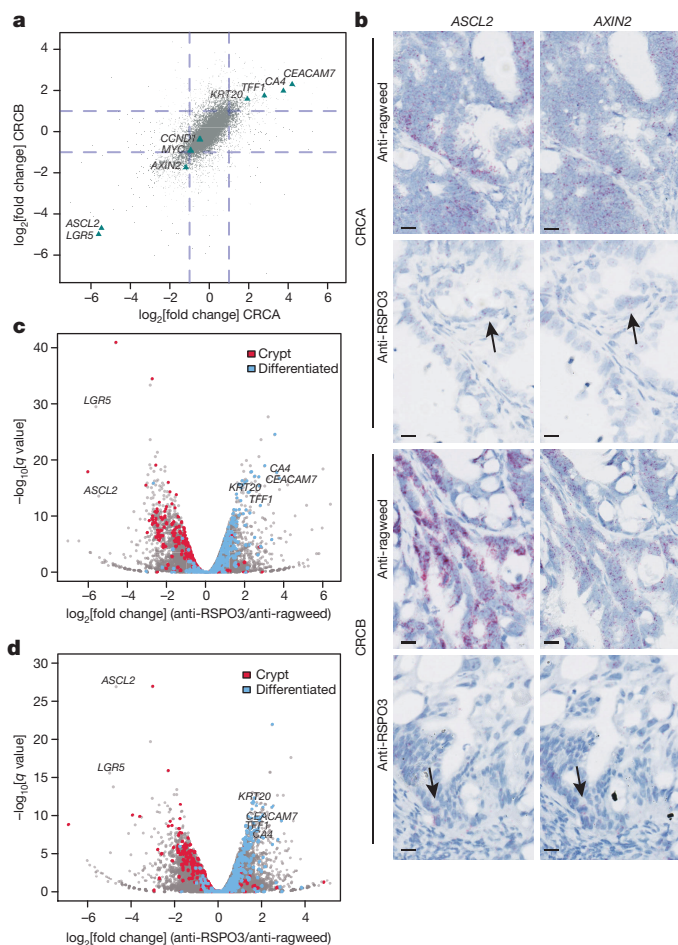
<sup>1</sup>Molecular Oncology, Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA. <sup>2</sup>Molecular Biology, Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA.

<sup>3</sup>Translational Oncology, Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA. <sup>4</sup>Antibody Engineering, Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA.

<sup>5</sup>Discovery Oncology, Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA. <sup>6</sup>Research Pathology, Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA.

<sup>7</sup>Protein Chemistry, Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, USA.



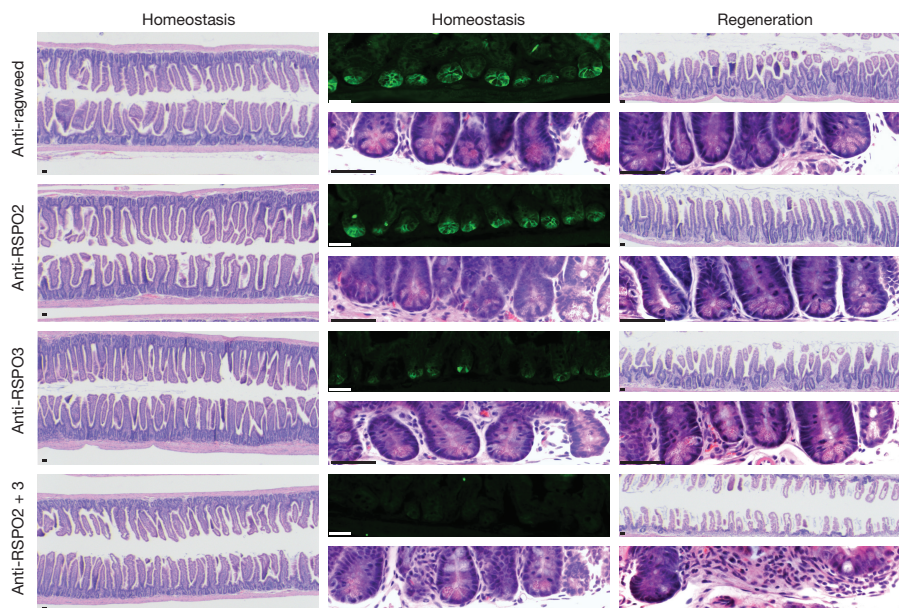


**Figure 2 | Anti-RSPO3 promotes differentiation.** **a**, Scatter plot of the expression response of CRCA and CRCB. Grey dots, individual genes. Triangles, specific genes. Dotted lines indicate a twofold change in expression. **b**, Representative images of *in situ* hybridization of *ASCL2* and *AXIN2*, as indicated. Arrows indicate residual expression. Scale bars, 20  $\mu\text{m}$ . **c**, **d**, Volcano plots indicating differentially expressed genes from CRCA (**c**) and CRCB (**d**). Genes enriched in the stem/undifferentiated compartment of the colon<sup>13</sup> are indicated in red. Genes enriched in the differentiated compartment of the colon<sup>13</sup> are indicated in blue.  $n = 3$  for all groups.

was more uniformly reduced in CRCB following anti-RSPO3 treatment (Extended Data Fig. 3). To extend these analyses, we generated RNA-sequencing data from day 4 samples (Supplementary Tables 1 and 2). Comparison of the global response of the two models revealed notable similarities (Fig. 2a), including the downregulation of Wnt target genes, which were statistically enriched in our data set (Extended Data Fig. 4). However, the effect on individual Wnt target genes varied. For example, *AXIN2*, *MYC* and *CCND1* were modestly reduced (Fig. 2a and Extended Data Fig. 4) and were not present in the top 100 most downregulated genes among those statistically significantly regulated (Supplementary Table 1). In contrast, *LGR5* ranked second and *ASCL2* ranked among the top five in both models (Fig. 2a, Extended Data Figs 4 and 5 and Supplementary Table 1). *In situ* hybridization confirmed the sensitivity of *ASCL2*, which was expressed in most of the undifferentiated cells of the tumours and was markedly reduced following anti-RSPO3 treatment (Fig. 2b). In addition to *LGR5* and *ASCL2*, which are well-characterized markers of intestinal stem cells<sup>9</sup>, the stem-cell marker genes *LRIG1* and *TERT* were also downregulated (Supplementary Table 1). The sensitivity of Wnt target genes expressed in stem cells suggest they require RSPO3 to achieve the high level of Wnt signalling that is associated with stem-cell activity<sup>10</sup>. Indeed, it was recently reported that in murine organoids *Ascl2* is expressed when a threshold of Wnt signalling is reached that is promoted by R-spondins<sup>11</sup>.

Genes that were upregulated upon anti-RSPO3 treatment included markers of differentiation<sup>12</sup> (Fig. 2a, c, d, Extended Data Fig. 5 and Supplementary Table 1), suggesting that anti-RSPO3 disrupts the cellular hierarchy within the tumour. Therefore, we compared the gene signatures of CRCA and CRCB with published gene signatures derived from the stem-cell and differentiated-cell compartments of the human colon<sup>13</sup>. In both models, treatment with anti-RSPO3 resulted in the downregulation of genes expressed in the stem-cell compartment and an upregulation of the genes expressed in differentiated cells (Fig. 2c, d and Extended Data Fig. 6). This effect was specific to RSPO3-fusion tumours, as anti-RSPO3 treatment had no effect on tumour growth or molecular response in two different *APC* mutant PDX models (Extended Data Fig. 7).

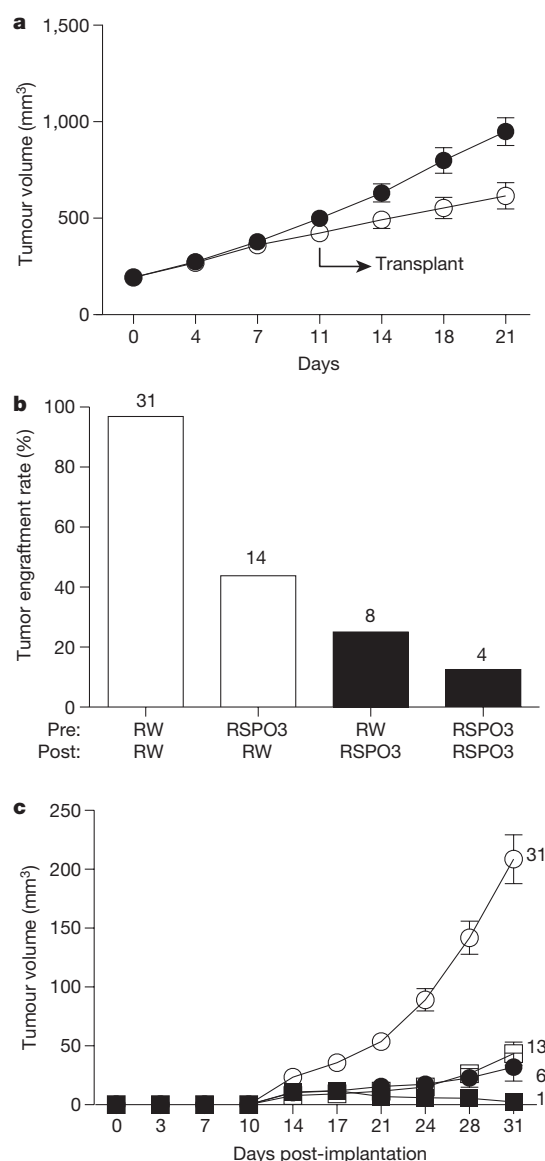
Our data indicate that anti-RSPO3 specifically shifts fusion tumours to a more differentiated phenotype, as previously reported in colon tumours following Wnt pathway inhibition<sup>14–16</sup>. To further explore this process, we characterized the molecular response at three different time points following a single dose of anti-RSPO3. *AXIN2*, *LGR5* and *ASCL2* were reduced within 24 h and *LGR5* and *ASCL2* remained low through



**Figure 3 | RSPO2 and RSPO3 are required for normal stem-cell function in the intestine.**

Representative sections of jejunum from animals treated with antibodies during homeostasis ( $n = 3$ ) or regeneration ( $n = 5$ ), stained with haematoxylin and eosin or anti-GFP. Scale bars, 50  $\mu\text{m}$ . Experiments on normal intestinal homeostasis were performed at least twice.





**Figure 4 | Anti-RSPO3 treatment reduces tumour propagation.**

**a**, Growth curves indicating when tumours were harvested to transplant into naive mice. Closed circles, anti-ragweed. Open circles, anti-RSPO3. **b**, Engraftment frequency of transplanted tumours from **a**. The number of tumours measured at two consecutive time points is indicated. RW, anti-ragweed; RSPO3, anti-RSPO3. **c**, Tumour growth following transplantation. Open circles, anti-ragweed-treated transplanted to anti-ragweed-treated. Open squares, anti-RSPO3-treated transplanted to anti-ragweed-treated. Closed circles, anti-ragweed-treated transplanted to anti-RSPO3-treated. Closed squares, anti-RSPO3-treated transplanted to anti-RSPO3-treated. The number of tumours at the end of the study is indicated. Data are means  $\pm$  s.e.m.

10 days. By contrast, markers of differentiation began to increase by 4 days and continued to increase through 10 days (Extended Data Fig. 8).

The immediate and robust effect of anti-RSPO3 on stem-cell markers, combined with reports indicating that R-spondin overexpression expands the intestinal stem-cell compartment<sup>17</sup> and is required for culture of intestinal stem cells<sup>18</sup>, led us to hypothesize that RSPO3 regulates stem-cell function. To investigate this possibility, we turned to the normal intestine, where stem-cell biology is well characterized<sup>9</sup>. *Lgr5* marks stem cells in the normal intestine that are sensitive to R-spondin overexpression<sup>17,19</sup>. While ablation of *Lgr5*-positive cells does not have an acute effect on homeostasis, these cells are required for regeneration following irradiation injury<sup>20,21</sup>. We therefore examined

the effect of blocking R-spondins in the intestine during homeostasis and regeneration. As the normal intestine expresses both *RSPO2* and *RSPO3* (Extended Data Fig. 2), we included an anti-*RSPO2*/anti-*RSPO3* combination group in the study. Similar to the effect of *Lgr5*<sup>+</sup> cell ablation<sup>20,21</sup>, there was minimal impact of antibody treatment during homeostasis, but a profound impact on regeneration following irradiation (Fig. 3 and Extended Data Fig. 9). Strong effects were observed in the combination group where *Lgr5*<sup>+</sup> cells were difficult to identify and the regeneration of the epithelium following irradiation was markedly impaired (Fig. 3). The severity of the phenotype of the combination group suggests there is functional redundancy between *Rspo2* and *Rspo3* in the normal intestine not present in *RSPO3*-fusion-positive tumours.

To address whether anti-*RSPO3* reduces stem-cell function in tumours, we performed a serial transplantation study. Mice implanted with CRCA tumours were treated with either control or anti-*RSPO3* antibody (Fig. 4a) and tumour fragments from each group were transplanted into naive recipient mice, which were subsequently treated with either control or anti-*RSPO3* antibody. Transplanted tumours from control animals engrafted at a 97% (31/32) rate and initiated robust tumour growth following transplantation (Fig. 4b, c). By contrast, the engraftment rate was only 44% (14/31) from anti-*RSPO3*-treated tumour fragments and these tumours grew more slowly (Fig. 4b, c). Continuation of anti-*RSPO3* treatment following transplantation profoundly impacted tumour engraftment and growth, with regression occasionally being observed in both groups (2/8 control pre-treated, 3/4 anti-*RSPO3* pre-treated) (Fig. 4b, c). To test whether the reduction in tumour engraftment was accompanied by a decrease in tumour-initiating cell content, we performed a flow cytometric analysis using CD133 and CD44, two proteins previously identified as markers for tumour-initiating cells in colon cancer<sup>22,23</sup>. Anti-*RSPO3* reduced the number of CD133<sup>+</sup> and CD44<sup>+</sup> cells in both models (Extended Data Fig. 10).

Cells with stem-cell properties have been identified in multiple tumour types and have formed the basis of a 'cancer stem cell hypothesis' where a population of cells within a tumour has enhanced long-term tumour-propagating potential and can promote relapse following therapy cessation<sup>22–26</sup>. It remains to be established whether a stem-cell compartment could drive colorectal tumour growth and whether targeting these cells could provide an effective therapeutic strategy. Our data, along with published data on the role of RSPOs in stem-cell biology<sup>17</sup> highlight a role for RSPO3 in regulating stem-cell function in *PTPRK(e1)*-*RSPO3(e2)*-fusion-positive tumours. *RSPO2* may have a similar role in tumours where fusions have been identified<sup>4,27</sup>. However, the role for *RSPO2* as a driver could be more complex as it has been proposed to be a tumour suppressor in other contexts<sup>28</sup>.

Molecular characterization of CRC has revealed that virtually all tumours exhibit aberrant Wnt signalling<sup>1</sup>. Wnt is a known regulator of stem-cell hierarchy with the highest level of activity associated with stem-cell properties and tumour-initiating ability<sup>9–11</sup>. Anti-*RSPO3* reduced, but did not eliminate Wnt activity, consistent with being a ligand-dependent amplifier of pathway activation<sup>6–8</sup>. This reduction was associated with tumour growth inhibition, suggesting that *RSPO3*-fusion tumours depend on the high level of Wnt activity associated with stem cells and promoted by R-spondins<sup>11</sup>. Correspondingly, direct inhibition of Wnt ligand production was recently reported to promote differentiation in *RSPO3*-fusion tumour models<sup>15</sup>. However, our data in the normal intestine suggest that inhibiting Wnt signalling during regeneration may not be tolerated (Fig. 3). The functional redundancy of the R-spondins present in the normal intestine creates a unique opportunity to specifically target the oncogenic driver of *RSPO3*-fusion tumours with fewer safety concerns.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 March; accepted 18 November 2015.

Published online 23 December 2015.

1. The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
2. De Sousa E Melo, F. *et al.* Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. *Nature Med.* **19**, 614–618 (2013).
3. Sadanandam, A. *et al.* Reconciliation of classification systems defining molecular subtypes of colorectal cancer: interrelationships and clinical implications. *Cell Cycle* **13**, 353–357 (2014).
4. Seshagiri, S. *et al.* Recurrent R-spondin fusions in colon cancer. *Nature* **488**, 660–664 (2012).
5. Siegel, R., Desantis, C. & Jemal, A. Colorectal cancer statistics, 2014. *CA Cancer J. Clin.* **64**, 104–117 (2014).
6. Kim, K. A. *et al.* R-spondin proteins: a novel link to beta-catenin activation. *Cell Cycle* **5**, 23–26 (2006).
7. de Lau, W. *et al.* Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling. *Nature* **476**, 293–297 (2011).
8. Hao, H. X. *et al.* ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner. *Nature* **485**, 195–200 (2012).
9. Clevers, H. The intestinal crypt, a prototype stem cell compartment. *Cell* **154**, 274–284 (2013).
10. Vermeulen, L. *et al.* Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. *Nature Cell Biol.* **12**, 468–476 (2010).
11. Schuijers, J. *et al.* Ascl2 acts as an R-spondin/Wnt-responsive switch to control stemness in intestinal crypts. *Cell Stem Cell* **16**, 158–170 (2015).
12. Jung, P. *et al.* Isolation and *in vitro* expansion of human colonic stem cells. *Nature Med.* **17**, 1225–1227 (2011).
13. Kosinski, C. *et al.* Gene expression patterns of human colon tops and basal crypts and BMP antagonists as intestinal stem cell niche factors. *Proc. Natl Acad. Sci. USA* **104**, 15418–15423 (2007).
14. Dow, L. E. *et al.* Apc restoration promotes cellular differentiation and reestablishes crypt homeostasis in colorectal cancer. *Cell* **161**, 1539–1552 (2015).
15. Madan, B. *et al.* Wnt addiction of genetically defined cancers reversed by PORCN inhibition. *Oncogene* (2015).
16. Scholer-Dahirel, A. *et al.* Maintenance of adenomatous polyposis coli (APC)-mutant colorectal cancer is dependent on Wnt/beta-catenin signaling. *Proc. Natl Acad. Sci. USA* **108**, 17135–17140 (2011).
17. Yan, K. S. *et al.* The intestinal stem cell markers Bmi1 and Lgr5 identify two functionally distinct populations. *Proc. Natl Acad. Sci. USA* **109**, 466–471 (2012).
18. Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures *in vitro* without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
19. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449**, 1003–1007 (2007).
20. Metcalfe, C., Kljavin, N. M., Ybarra, R. & de Sauvage, F. J. *Lgr5*<sup>+</sup> stem cells are indispensable for radiation-induced intestinal regeneration. *Cell Stem Cell* **14**, 149–159 (2014).
21. Tian, H. *et al.* A reserve stem cell population in small intestine renders *Lgr5*-positive cells dispensable. *Nature* **478**, 255–259 (2011).
22. O'Brien, C. A., Pollett, A., Gallinger, S. & Dick, J. E. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* **445**, 106–110 (2007).
23. Ricci-Vitiani, L. *et al.* Identification and expansion of human colon-cancer-initiating cells. *Nature* **445**, 111–115 (2007).
24. Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. *Proc. Natl Acad. Sci. USA* **100**, 3983–3988 (2003).
25. Bonnet, D. & Dick, J. E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature Med.* **3**, 730–737 (1997).
26. Kreso, A. & Dick, J. E. Evolution of the cancer stem cell model. *Cell Stem Cell* **14**, 275–291 (2014).
27. Robinson, D. *et al.* Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215–1228 (2015).
28. Wu, C. *et al.* RSP02–LGR5 signaling has tumour-suppressive activity in colorectal cancer. *Nature Commun.* **5**, 3149 (2014).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** We would like to thank D. Jakubiak, Z. Modrusan, J. Stinson, P. Havery and the Sequencing and Microarray laboratories for assistance with the animal studies and RNA-seq data generation.

**Author Contributions** E.E.S. collected and analysed data from antibody characterization, tumour response characterization, normal intestinal response experiments, and wrote the manuscript. S.D. analysed the RNA-seq data and performed the GSEA. F.d.S.e.M. assisted in the analysis of the RNA-seq and flow cytometry data. J.T. and E.B. assisted in the design and interpretation of efficacy studies. N.K. assisted in the irradiation experiments. C.T., C.C., J.-A.H. and Y.W. generated and assisted in the characterization of the antibodies. T.B. purified proteins and antibodies. R.F. assisted in the histopathology. C.M. assisted with normal homeostasis experiments. E.W.S. analysed exome sequencing data. C.M., X.Y. and B.Y. screened samples to identify tumour models. T.P. generated RNA *in situ* hybridization data. F.d.S. was involved in study design, interpretation and manuscript preparation.

**Author Information** Sequence data have been deposited at the European Genome-phenome Archive (EGA; <https://www.ebi.ac.uk/ega/home>) under accession number EGAS00001001462. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.d.S. ([desauvage.fred@gene.com](mailto:desauvage.fred@gene.com)).

## METHODS

No statistical methods were used to predetermine sample size.

**Antibody affinity.** Binding affinities of anti-RSPO antibodies were measured by surface plasmon resonance (SPR) using a BIAcore™-2000 instrument. The CM5 biosensor chip was activated with *N*-ethyl-*N'*-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) and *N*-hydroxysuccinimide (NHS) reagents according to the supplier's (GE Healthcare Biosciences) instructions. RSPO antigens (human RSPO2 and human RSPO3 (Genentech, Inc.); human RSPO1, human RSPO4, and mouse RSPOs (R&D Systems) were immobilized onto the biosensor chip to achieve approximately 30 response units (RU), followed by blocking with 1 M ethanolamine. For kinetic measurements, fourfold serial dilutions of anti-RSPO antibodies were injected in a range of 0.0976–100 nM HBS-P buffer (0.01 M HEPES pH 7.4, 0.15 M NaCl, 0.005% surfactant P20) at 25 °C with a flow rate of 30  $\mu$ l min<sup>-1</sup>. Association rates ( $k_{on}$ ) and dissociation rates ( $k_{off}$ ) were calculated using a simple one-to-one Langmuir binding model (BIAcore Evaluation T200 Software version 2.0). The equilibrium dissociation constant ( $K_d$ ) was calculated as the ratio  $k_{off}/k_{on}$ .

**Antibody activity.** 293T cells (Genentech cell bank, authenticated by STR profiling, SNP fingerprinting and mycoplasma tested) were reverse transfected and plated in 100  $\mu$ l of DMEM containing 2.5% fetal bovine serum under the following conditions per well: 9,000 cells, 0.04  $\mu$ g Topbrite 25 plasmid, 0.02  $\mu$ g SV-40 Renilla plasmid, 0.25  $\mu$ l Eugene 6 (Promega). Following 16–20 h of culture, cells were stimulated with 25  $\mu$ l of a 5  $\times$  Wnt/RSPO solution in DMEM 10% fetal bovine serum. Cells were stimulated with a final concentration of 10 ng ml<sup>-1</sup> rmWnt3a (R&D Systems) and the EC<sub>50</sub> of RSPO2 (11.6 pM, Genentech, Inc.), or the EC<sub>50</sub> of RSPO3 (10.5 pM, Genentech, Inc.)  $\pm$  anti-RSPO2 or anti-RSPO3 for 6 h at 37 °C. Following, luciferase activity was detected using the Promega Dual-Glo system (Promega) according to the manufacturer's instructions. Data were analysed as a ratio of Firefly/Renilla (RLU WNT reporter).

**Animal studies.** All studies involving animals were approved by Genentech's Institutional Animal Care and Use Committee and adhere to the NRC Guidelines for the Care and Use of Laboratory Animals. For xenograft studies, animals were randomized into treatment groups based on starting tumour volumes, when tumours reached a mean volume of approximately 150–300 mm<sup>3</sup>. Animals were humanely euthanized according to the following criteria: clinical signs of persistent distress or pain, significant body weight loss (>20%), tumour size exceeding 2,000 mm<sup>3</sup>, or when tumours ulcerate. Maximum tumour size permitted by the Institutional Animal Care and Use Committee is 3,000 mm<sup>3</sup> and in none of the experiments was this limit exceeded.

**Patient-derived xenograft studies.** PDX studies were performed by implanting primary tumour fragments subcutaneously in Balb/C nude mice. Animals were distributed into treatment groups ( $n = 10$  per group) when tumours reached a mean volume of approximately 150–300 mm<sup>3</sup>. Control antibody (anti-ragweed; 30 mg kg<sup>-1</sup>) or anti-RSPO3 (30 mg kg<sup>-1</sup>) was injected in the intraperitoneal space twice a week for 3–4 weeks. Tumour size and body weight measurements were collected twice a week. Tumour volume was determined using digital callipers (Fred V. Fowler Company, Inc.) using the formula  $l \times w^2$ , where  $l$  is length and  $w$  is width. Mice with tumour volumes exceeding 2,000 mm<sup>3</sup>, or that had lost  $\geq 20\%$  of their initial bodyweight were humanely euthanized. Experiments were double-blinded.

**Homeostasis experiments.** The effects of antibodies on intestinal stem cells during homeostasis was explored using *LGR5-DTR* mice<sup>21</sup> ( $n = 3$  per group). Males and females 12–16 weeks were used. Animals were injected with antibody (30 mg kg<sup>-1</sup>) twice a week and samples of intestine were collected on day 4 or day 9.

**Regeneration experiments.** Regeneration experiments were performed on 6–12 week Balb/c/J female mice ( $n = 5$  per group). Mice were injected with antibodies (30 mg kg<sup>-1</sup>) on days 0, 3 and 6 and subjected to 10 Gy whole-body irradiation on day 3. On day 7, 4 days following irradiation, the mice were humanely euthanized and samples of intestine collected for histology. Homeostasis and regeneration experiments were not blinded.

**Histology.** Samples were formalin-fixed and paraffin-embedded using standard procedures. For Ki67 antibody stains, rehydrated sections were pressure-cooked for 15 min in antigen unmasking buffer (DAKO), blocked in serum-free protein block (DAKO) and incubated in anti-Ki67 (1:400, Sigma) overnight. Sections were then incubated in HRP-conjugated anti-rabbit (DAKO) and detected with DAB reaction (DAKO). GFP in *LGR5-DTR* mice was detected in frozen sections with anti-GFP (Torrey Pines) followed by Alexa Fluor 488 (Life Technologies).

**Tumour dissociation and flow cytometry.** Tumour-bearing mice were treated with anti-ragweed or anti-RSPO3 at 30 mg kg<sup>-1</sup> for 14 days. Tumours were collected from  $n = 5$  mice and digested enzymatically for 30 min with a mixture of collagenase (1.5 mg ml<sup>-1</sup>; Roche) and hyaluronidase (20  $\mu$ g ml<sup>-1</sup>) at 37 °C. The cells were then filtered (70  $\mu$ m pore size) and washed. Cell debris was removed by lympholyte (M; Cedarlane) centrifugation or ACK lysis. The cell suspensions were stained with the following antibodies for flow cytometry analysis: AC133 (Miltenyi

Biotec, 1:100), CD44 (BD Biosciences, 1:100). Dead cells were excluded with Sytox blue (1:1,000, Invitrogen). Gating of singlet live cells was used to determine the percentage of CD133 or CD44 within each independent tumour.

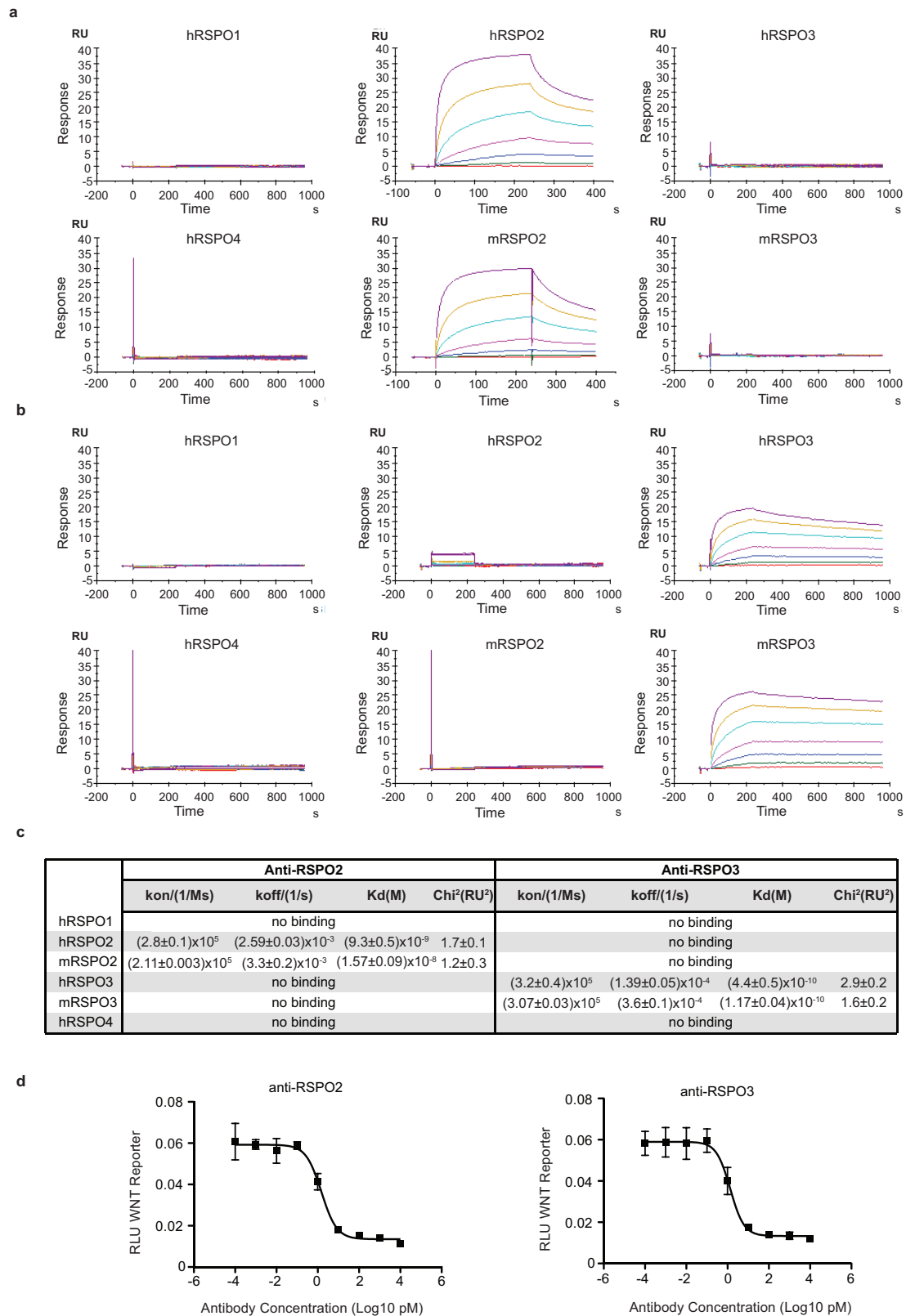
**RNA isolation and PCR.** RNA from samples was isolated using a RNeasy Plus kit according to the manufacturer's instructions (Qiagen). qRT-PCR was performed in 10  $\mu$ l reactions using 50 ng total RNA using One-step Real-time RT-PCR mastermix (Life Technologies) according to the manufacturer's instructions. The following Taqman assays from Life Technologies were used: AXIN2 (Hs00610244\_m1), ASCL2 (Hs00270888\_s1), LGR5 (Hs00969422\_m1), CA4 (Hs00426343\_m1), MYC (Hs00153408\_m1), CEACAM7 (Hs03988977\_m1). Identification of RSPO3 fusions was performed using a *PTPRK(e1)-RSPO3(e2)*, specific Taqman assay: forward (TCTCCTTGGCCTCTCCTGGGAT), reverse (TTGGCAGCCTTGACTAACGTT), probe (TTCTCCGCACTGCATC).

**RNA-seq.** RNA-seq libraries were prepared using TruSeq RNA Sample Preparation kit (Illumina, CA). The libraries were sequenced on Illumina HiSeq 2500 sequencers and we obtained on average 34 million single-end reads (50 bp) per sample. RNA-seq reads were aligned to the human genome version NCBI GRCh37 using GSNAP. Expression counts per gene were obtained by counting the number of reads aligned concordantly within a pair and uniquely to each gene locus as defined by NCBI, Ensembl gene annotations, and RefSeq mRNA sequences. Differential gene expression analysis was performed using edgeR. Gene enrichment analysis was performed on the edgeR differential expression results using the GSEA preranked tool available through the Broad's GSEA application. The colon crypt gene sets used for GSEA analysis were derived from the paper by Kosinsky *et al.*<sup>13</sup>. Wnt target gene sets were derived from the paper of Fevr *et al.*<sup>29</sup>. DESeq was used to compute the variance stabilized expression values for plotting the expression heat maps.

**In situ hybridization.** Non-isotopic *in situ* hybridization was performed using probes from Affymetrix to human ASCL2 (catalogue no. VA1-17147), and AXIN2 (no. VA1-10388) with a probe set to *Bacillus subtilis* dihydropicolinate reductase (no. VF-11712) used as a negative control. A modified, manual version of the ViewRNA eZ Detection Kit (Affymetrix) was used for amplification and detection of hybridized signal. In brief, 4- $\mu$ m formalin-fixed paraffin-embedded sections were deparaffinized in Xylenes (Richard Allen Scientific) and then air-dried. Sample pre-treatment was done by incubating slides in 1  $\times$  Affymetrix Pretreatment Solution for 15 min at 99 °C in a PT Module (Thermo Fisher) then digested in Protease QF (Affymetrix) diluted 1:120 in PBS for 25 min at 40 °C. Slides were then post-fixed in 4% paraformaldehyde and blocked with Bloxall (Vector Labs) before hybridization with probes diluted 1:65 for 2.5 h at 40 °C. Branched DNA amplification was done by incubating in Amp1 solution for 25 min, Amp2 solution for 18 min and Amp3 solution for 18 min all at 41.7 °C then Label Probe-AP solution at 1:1,000 dilution for 15 min at 40 °C. Hybridization was visualized using Warp Red substrate (Biocare) and samples were counterstained in haematoxylin. Slides were imaged using a Nanozoomer XR (Hamamatsu).

**Exome sequencing.** Sequencing reads were mapped to a combined genome with both the UCSC human genome (GRCh37/hg19) and the mouse genome (mm9) using BWA software<sup>30</sup> set to default parameters. Because of potential mouse stromal contamination, using a combined genome has previously been shown to increase the overall variant calling accuracy when analysing xenograft exome data<sup>31</sup>. Local realignment, duplicate marking were performed as described previously<sup>32</sup>. Reads unambiguously mapped to the human genome were used for variant calling using the Strelka<sup>33</sup> program. Known germline variants represented in dbSNP Build 131 (ref. 34) or 6,515 previously published normal exomes<sup>35</sup>, but not represented in COSMIC v62 (ref. 36), were removed.

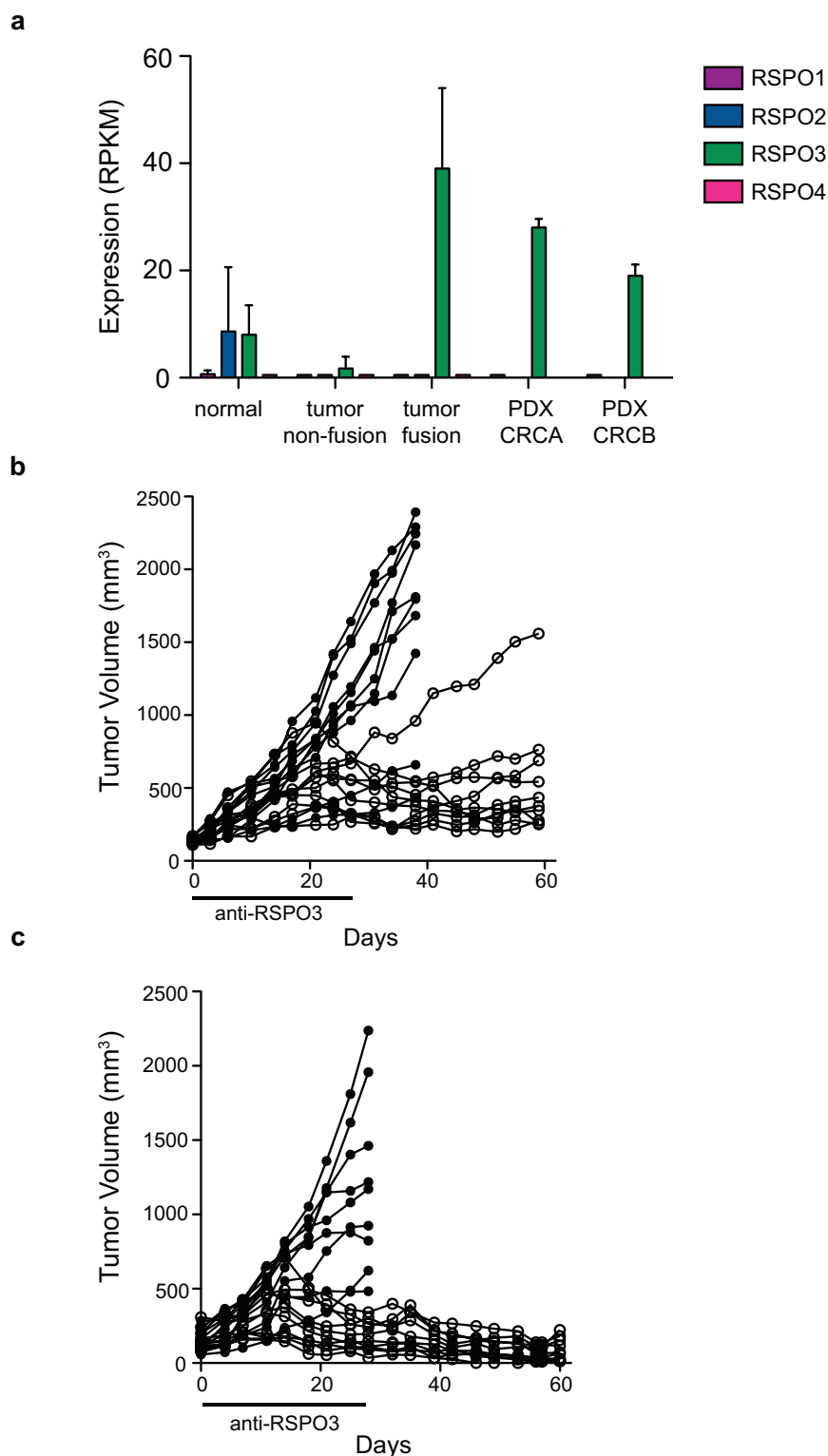
29. Fevr, T., Robine, S., Louvard, D. & Huelsen, J. Wnt/ $\beta$ -catenin is essential for intestinal homeostasis and maintenance of intestinal stem cells. *Mol. Cell. Biol.* **27**, 7551–7559 (2007).
30. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
31. Tso, K. Y., Lee, S. D., Lo, K. W. & Yip, K. Y. Are special read alignment strategies necessary and cost-effective when handling sequencing reads from patient-derived tumor xenografts? *BMC Genomics* **15**, 1172 (2014).
32. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* **43**, 491–498 (2011).
33. Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).
34. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
35. Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**, 216–220 (2013).
36. Forbes, S. A. *et al.* COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res.* **38**, D652–D657 (2010).



**Extended Data Figure 1 | Characteristics of anti-RSPO2 and anti-RSPO3 antibodies.** **a**, Surface plasmon resonance sensograms of anti-RSPO2 (0–125 nM) binding to human and mouse R-spondins (as indicated; h, human; m, mouse). **b**, Surface plasmon resonance sensograms of anti-RSPO3 (0–125 nM) binding to human and mouse R-spondins (as indicated). **c**, Affinities of anti-RSPO2 and anti-RSPO3 binding to human

and mouse R-spondins as indicated.  $K_d$ , equilibrium dissociation constant calculated as the ratio of  $k_{off}/k_{on}$ . Data are means  $\pm$  s.d. **d**, Functional blocking of Wnt reporter assay stimulated by 11.6 pM human RSPO2 or 10.5 pM human RSPO3 and 10 ng ml<sup>-1</sup> rmWNT3a (as indicated). Data represent the mean  $\pm$  s.e.m. of three independent experiments.

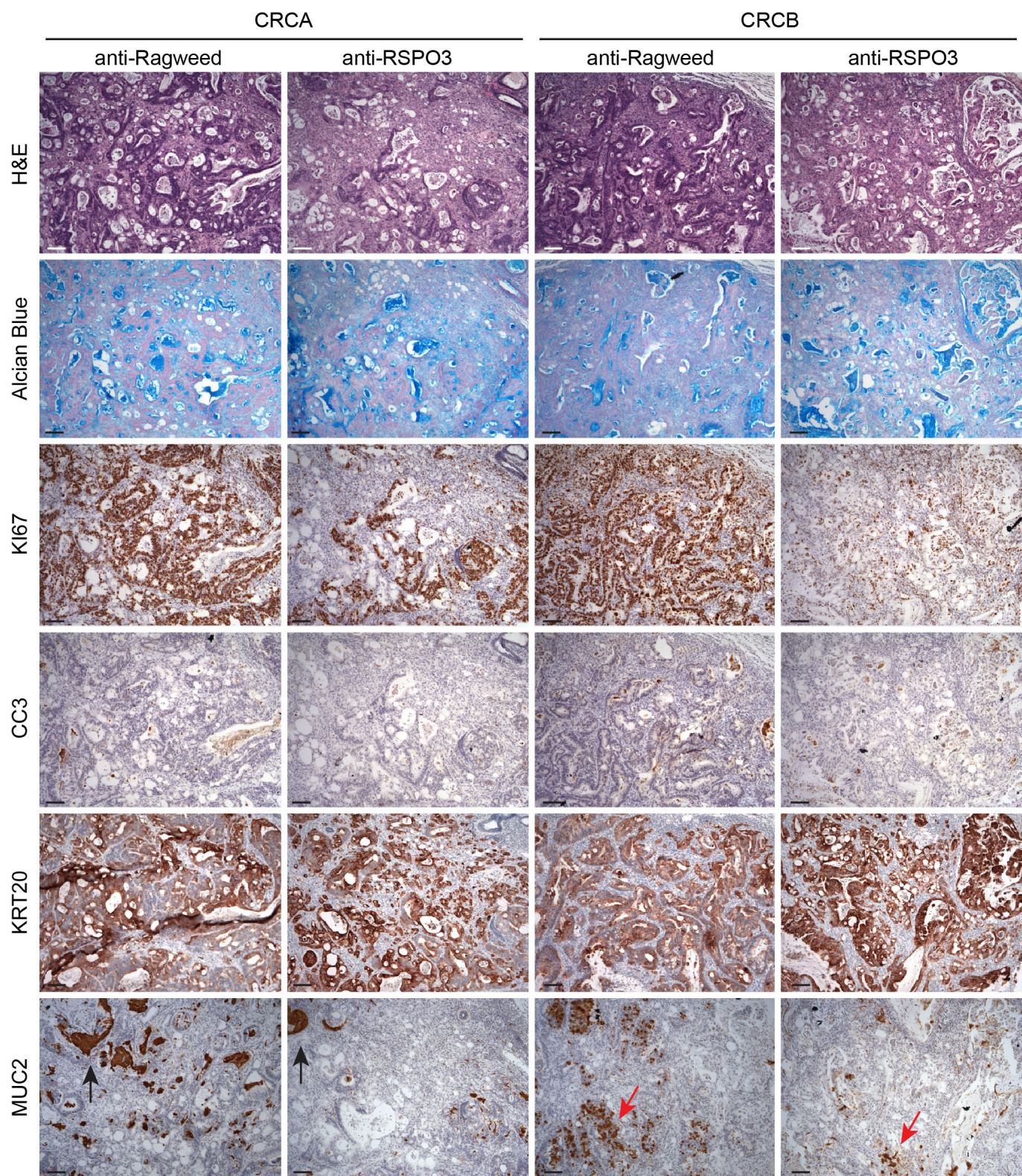




**Extended Data Figure 2 | Tumour growth inhibition persists following anti-RSPO3 treatment cessation.** **a**, RNA-seq expression of *RSPO1-4* and identification of *RSPO3*-fusion by fusion-specific quantitative reverse transcription PCR (qRT-PCR) in colon tumour, normal and PDX models. Expression data from normal colon and colon tumours were analysed from the publicly available data EGA accession EGAS00001000288 (ref. 4). Data are mean  $\pm$  s.d. Normal  $n = 72$ ; non-fusion tumour  $n = 67$ ; *PTPRK(e1)-RSPO3(e2)*  $n = 4$ ; CRCA  $n = 3$  different tumours from the model; CRCB  $n = 3$  different tumours from the model. RPKM, reads per kilobase per

million. Expression levels greater than 0 but less than 0.5 RPKM are graphed as 0.5. **b**, Individual tumour measurements from CRCA treated with anti-RSPO3 ( $30 \text{ mg kg}^{-1}$ ) twice a week for 4 weeks. Tumours were monitored for 30–40 days following the last dose of antibody ( $n = 10$ ). **c**, Individual tumour measurements of CRCB data presented in Fig. 1. Anti-ragweed,  $n = 9$ ; anti-RSPO3,  $n = 10$ . Tumour growth was followed for an additional 30–40 days following the last dose of antibody. Closed circles, anti-ragweed. Open circles, anti-RSPO3. Black bar indicates the time during which mice were treated with antibodies.

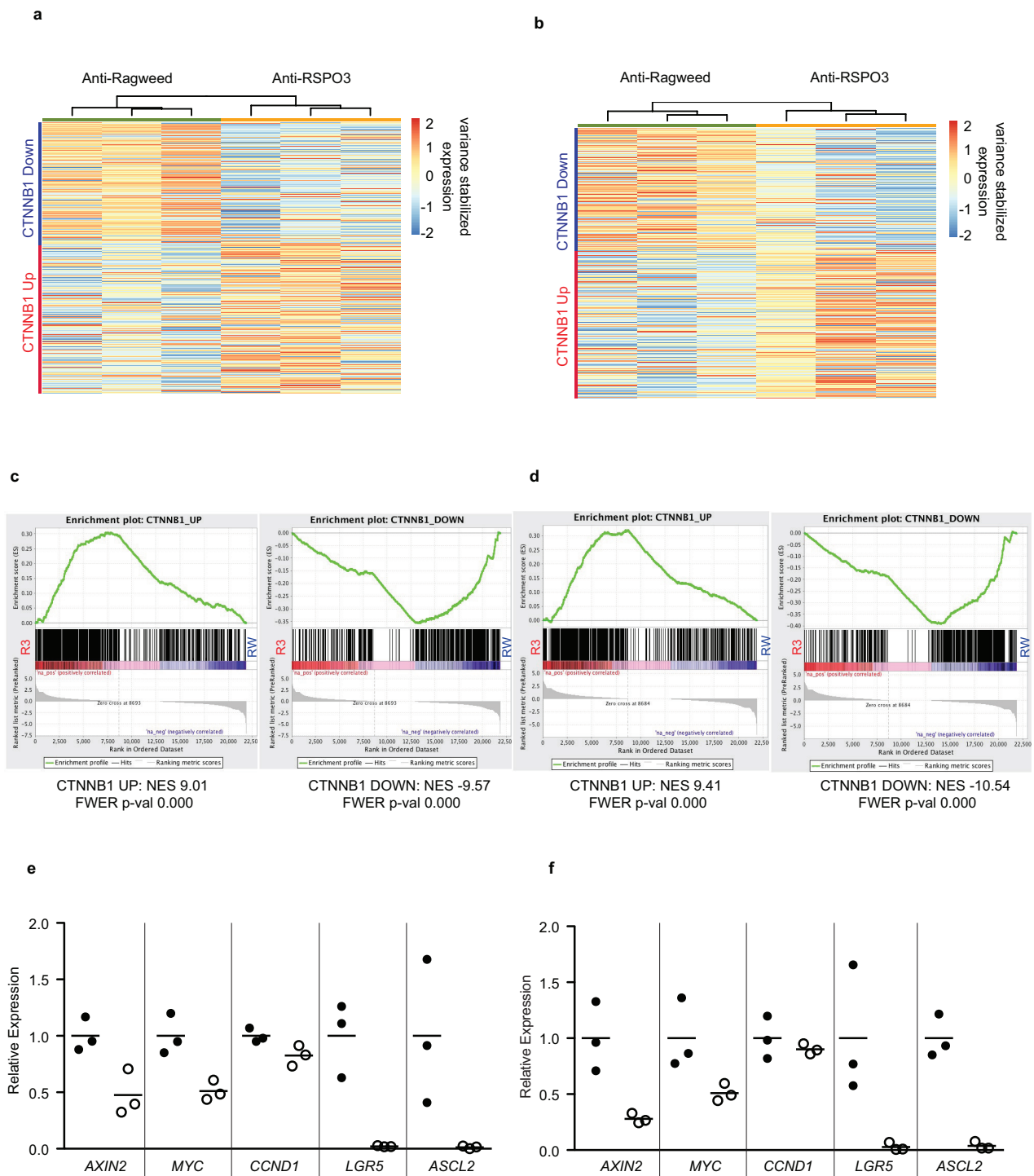




**Extended Data Figure 3 | Histopathological analysis of tumours from CRCA and CRCB.** Sections from CRCA and CRCB (representative tumour from  $n = 3$  analysed) collected 4 days following the initiation of

dosing and stained as indicated. CC3, cleaved caspase 3. Black arrows highlight extracellular MUC2 positivity, red arrows highlight intracellular MUC2 positivity. Scale bar, 100  $\mu\text{m}$ .

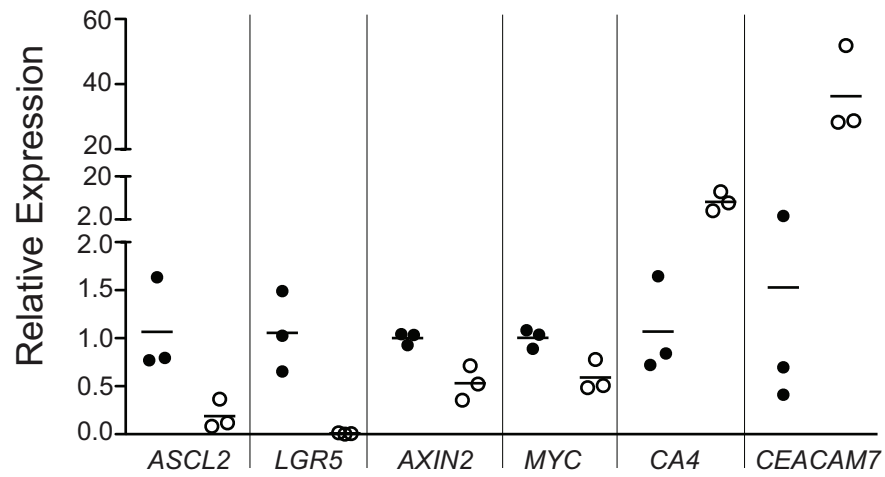




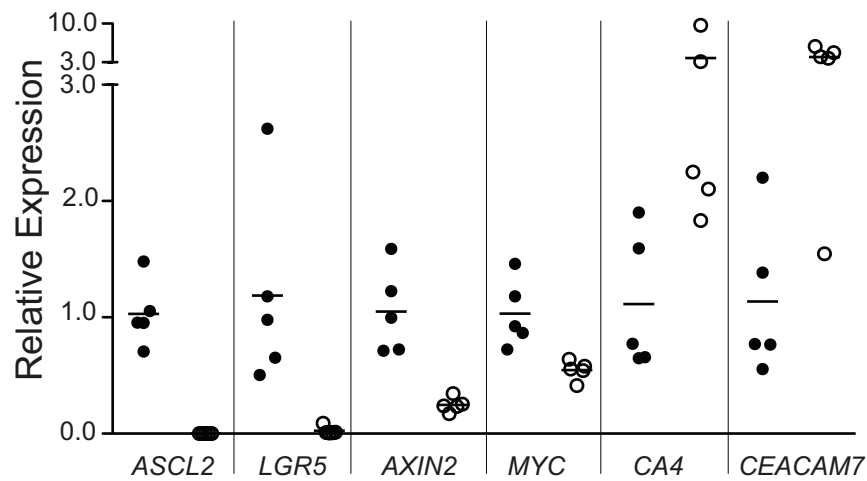
**Extended Data Figure 4 | Effects of anti-RSPO3 treatment on Wnt pathway activity.** **a**, Heat map of a CTNNB1 gene signature<sup>29</sup> in CRCA samples,  $n = 3$  tumours per group. **b**, Heat map of a CTNNB1 gene signature<sup>29</sup> in CRCB,  $n = 3$  tumours per group. **c**, Gene set enrichment analysis of CRCA using a CTNNB1 gene signature<sup>29</sup>. RW, anti-Ragweed-treated; R3, anti-RSPO3-treated; NES, normalized enrichment score;

FWER, family-wise error rate. **d**, Gene set enrichment analysis of CRCB using a CTNNB1 gene signature<sup>29</sup>. **e**, **f**, RNA-seq expression of a subset of Wnt target genes in CRCA (**e**) and CRCB (**f**). Data are normalized to control group. Horizontal line indicates the mean. Closed circles, anti-ragweed treated. Open circles, anti-RSPO3 treated.  $n = 3$  tumours.

**a**



**b**

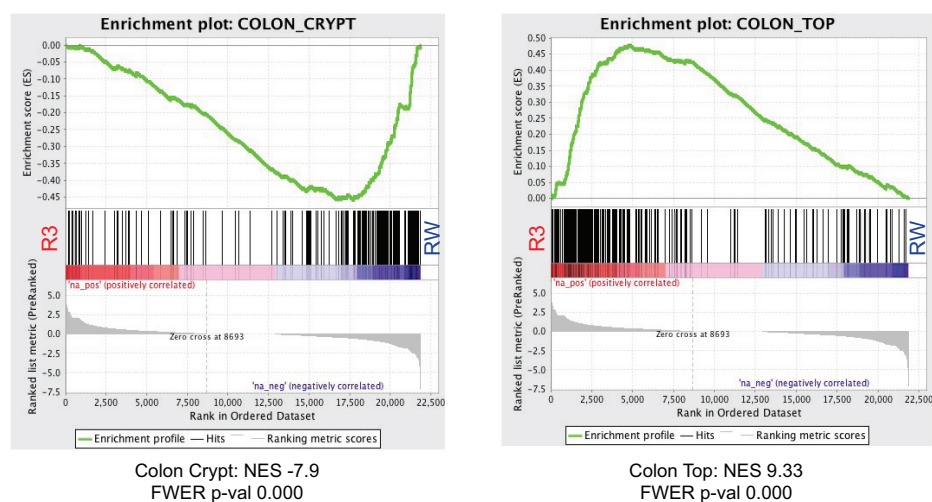


**Extended Data Figure 5 | Validation of gene expression changes by qRT-PCR. a, b,** Relative gene expression in CRCA and CRCB tumour samples as assessed by qRT-PCR. Closed circles, anti-ragweed-treated

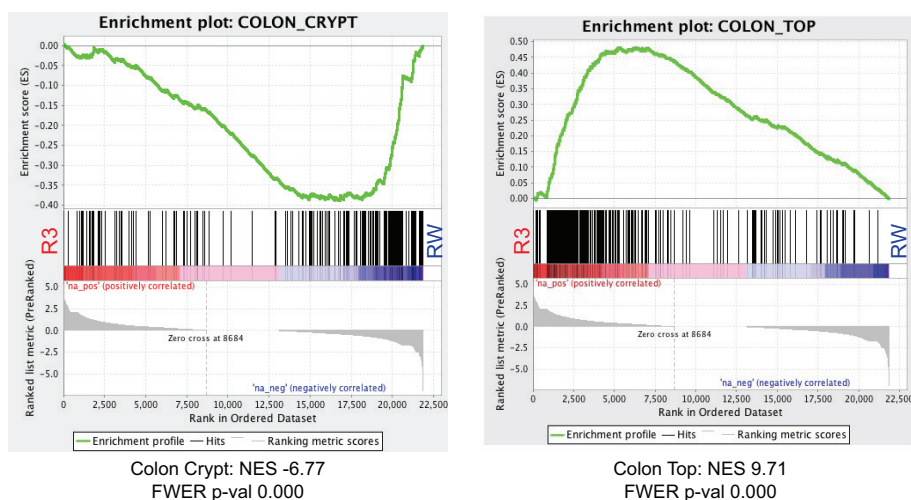
tumours. Open circles, anti-RSPO3-treated tumours. Horizontal line indicates the mean.  $n = 3$  for CRCA data and  $n = 4$  for CRCB. qRT-PCR data were validated from at least two independent experiments.



a

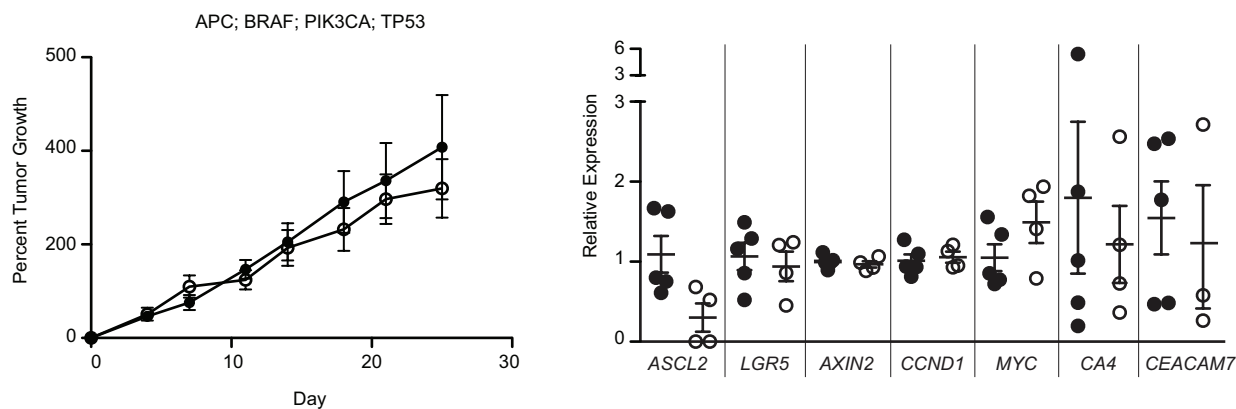


b

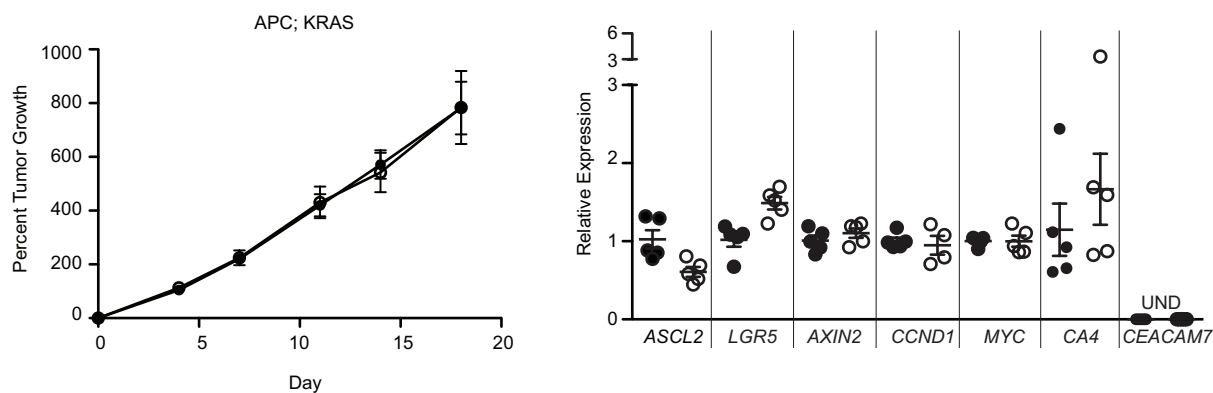


**Extended Data Figure 6 | Anti-RSPO3 treatment promotes differentiation.** **a**, Gene set enrichment using the Kosinsky *et al.* gene set<sup>13</sup> in CRCA. **b**, Gene set enrichment using the Kosinsky *et al.* gene set<sup>13</sup> in CRCB.  $n = 3$  per group. RW, anti-ragweed-treated; R3, anti-RSPO3-treated; NES, normalized enrichment score.

a

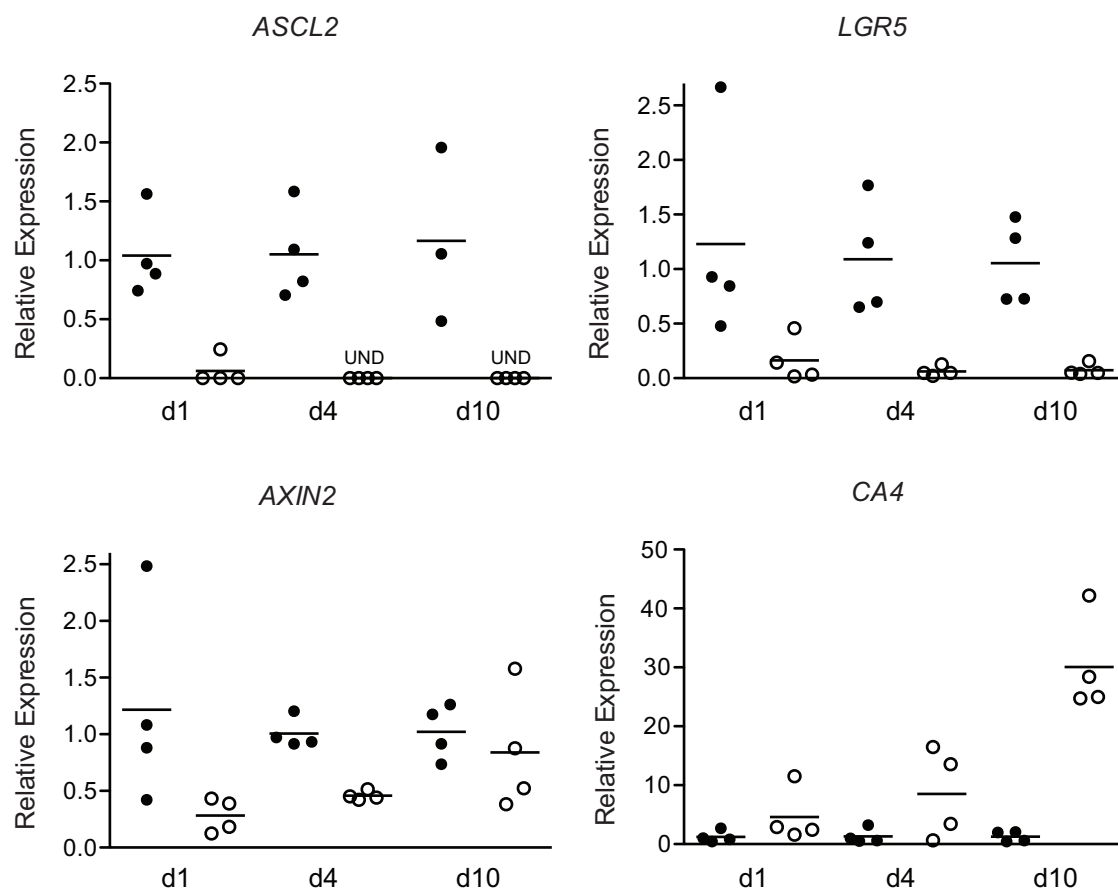


b



**Extended Data Figure 7 | APC mutant PDX models fail to respond to anti-RSPO3 treatment.** Tumour growth inhibition (left) and qRT-PCR of target genes (right) in APC mutant PDX models. **a**, APC(1554fs, 8fs, E941\*);BRAF(V600E);PIK3CA(E545K), TP53(R248W) mutant PDX model.  $n = 10$ , anti-ragweed treated,  $n = 9$ , anti-RSPO3-treated. **b**, APC(757fs, E1408\*);KRAS(G12D) model. Mice were treated

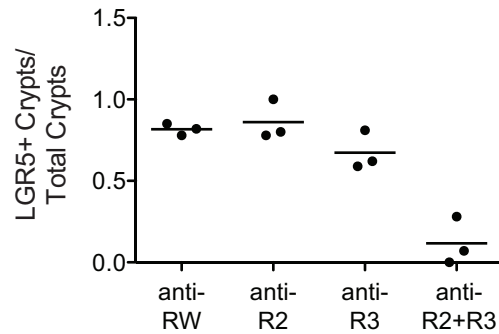
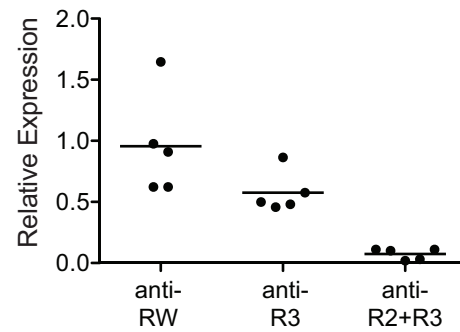
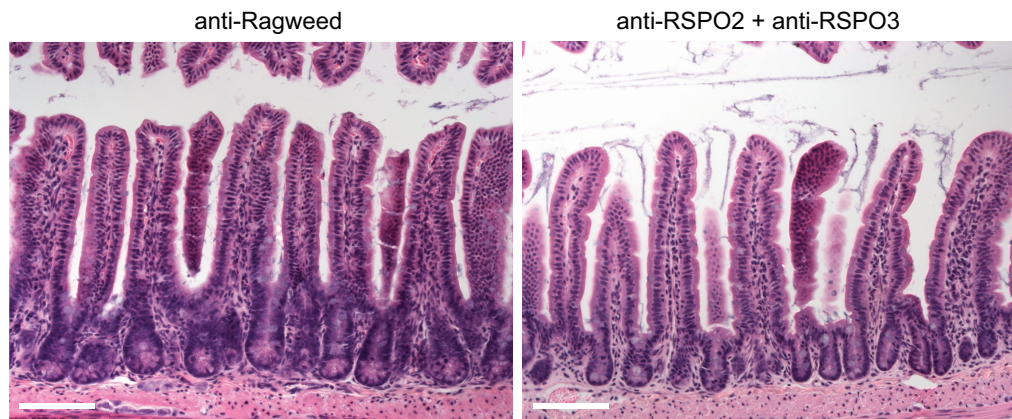
with antibodies ( $30 \text{ mg kg}^{-1}$ ) for 3–4 weeks ( $n = 10$  animals per group). Tumour growth data were normalized to initial tumour volume upon treatment ( $100\text{--}150 \text{ mm}^3$ ). qRT-PCR data ( $n = 5$ ) were collected at 8 (**a**) or 4 (**b**) days after the initiation of dosing. Data are plotted as mean  $\pm$  s.e.m. Closed circles, anti-ragweed. Open circles, anti-RSPO3. UND, undetermined  $C_t$  value.



**Extended Data Figure 8 | The downregulation of stem-cell marker gene expression is an early and robust response to anti-RSPO3 treatment.**

**a**, Relative gene expression in CRCA tumour samples collected 1, 4 and 10 days following a single injection of anti-RSPO3 ( $30 \text{ mg kg}^{-1}$ ).

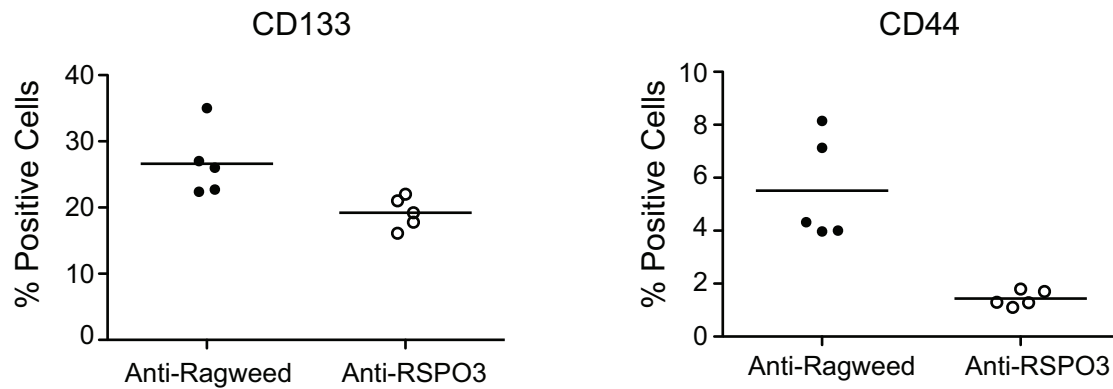
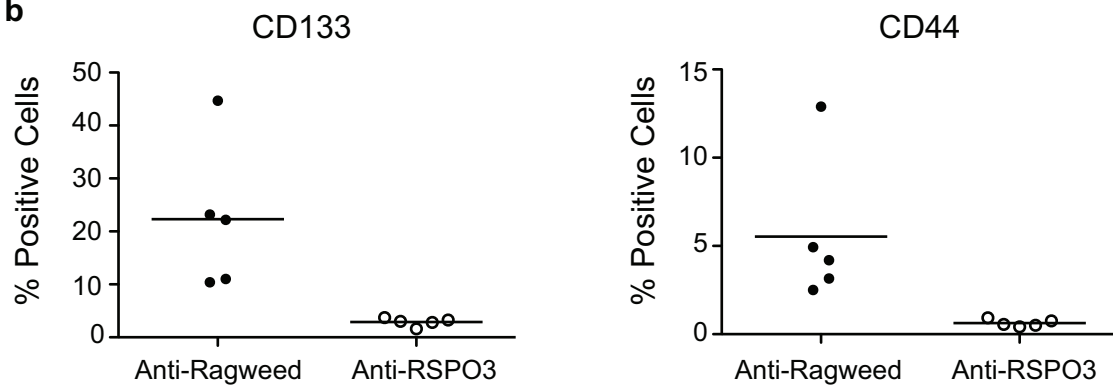
Closed circles, anti-ragweed-treated. Open circles, anti-RSPO3-treated. Horizontal line indicates the mean.  $n = 5$  tumours per group per time point. UND, undetermined  $C_t$  value.

**a****b****c**

**Extended Data Figure 9 | Combined treatment of anti-RSPO2 and anti-RSPO3 reduces *LGR5*.** **a**, Quantification of *Lgr5*<sup>+</sup> cells in the jejunum of mice treated with antibody (indicated on *x* axis) for 4 days. *Lgr5*-GFP-positive crypts/total crypts ( $n = 10$  for control,  $n = 20$  for antibody treated) were counted in three different mice. Horizontal line indicates the mean.

**b**, Relative expression assessed by qRT-PCR from RNA isolated from the ileum of animals treated with antibody for 9 days. Data are normalized to the control antibody treated group.  $n = 5$ , horizontal line indicates the mean. **c**, Haematoxylin and eosin stained sections from the intestine of mice treated with indicated antibodies for 9 days. Scale bar, 100  $\mu$ m.



**a****b**

**Extended Data Figure 10 | Anti-RSPO3 treatment reduces tumour-initiating cell content.** **a**, Percentage of CD133 (left) and CD44 (right)-positive cells as a percentage of live cells following 14 days of anti-RSPO3 treatment of CRCA. **b**, Percentage of CD133 (left) and CD44 (right)-

positive cells as a percentage of live cells following 8 days of anti-RSPO3 treatment of CRCB. Closed circles, anti-ragweed treated. Open circles, anti-RSPO3 treated. Horizontal line indicates the mean.  $n = 5$ .

# Species difference in ANP32A underlies influenza A virus polymerase host restriction

Jason S. Long<sup>1</sup>, Efstathios S. Giotis<sup>1</sup>, Olivier Moncorgé<sup>2</sup>, Rebecca Frise<sup>1</sup>, Bhakti Mistry<sup>1</sup>, Joe James<sup>1,3</sup>, Mireille Morisson<sup>4</sup>, Munir Iqbal<sup>3</sup>, Alain Vignal<sup>4</sup>, Michael A. Skinner<sup>1</sup> & Wendy S. Barclay<sup>1</sup>

**Influenza pandemics occur unpredictably when zoonotic influenza viruses with novel antigenicity acquire the ability to transmit amongst humans<sup>1</sup>. Host range breaches are limited by incompatibilities between avian virus components and the human host. Barriers include receptor preference, virion stability and poor activity of the avian virus RNA-dependent RNA polymerase in human cells<sup>2</sup>. Mutants of the heterotrimeric viral polymerase components, particularly PB2 protein, are selected during mammalian adaptation, but their mode of action is unknown<sup>3–6</sup>. We show that a species-specific difference in host protein ANP32A accounts for the suboptimal function of avian virus polymerase in mammalian cells. Avian ANP32A possesses an additional 33 amino acids between the leucine-rich repeats and carboxy-terminal low-complexity acidic region domains. In mammalian cells, avian ANP32A rescued the suboptimal function of avian virus polymerase to levels similar to mammalian-adapted polymerase. Deletion of the avian-specific sequence from chicken ANP32A abrogated this activity, whereas its insertion into human ANP32A, or closely related ANP32B, supported avian virus polymerase function. Substitutions, such as PB2(E627K), were rapidly selected upon infection of humans with avian H5N1 or H7N9 influenza viruses, adapting the viral polymerase for the shorter mammalian ANP32A. Thus ANP32A represents an essential host partner co-opted to support influenza virus replication and is a candidate host target for novel antivirals.**

The negative sense RNA genome of influenza virus is delivered into host cells associated with the viral RNA-dependent RNA polymerase, a heterotrimeric complex of PB1, PB2 and PA proteins and the viral nucleoprotein NP. The viral ribonucleoprotein (vRNP) complex traffics to the cell nucleus where, in complex with various co-opted host factors, the viral polymerase directs transcription and replication of the genome<sup>7</sup>. This intimate virus–host interaction is suboptimal for an avian influenza virus in mammalian cells, such that productive replication or onwards transmission do not occur until adaptive mutations are selected<sup>3,4,8,9</sup>. The host range restriction of influenza A virus polymerase can be recapitulated *in vitro*, whereby functional viral polymerase, reconstituted by expression of PB1, PB2, PA and NP, directs amplification and expression of a model viral-like RNA<sup>10</sup>. In mammalian cells, the low activity of reconstituted avian virus polymerase can be significantly increased by a single amino acid substitution (E627K) in PB2<sup>11,12</sup>. In contrast, in avian cells, polymerases bearing either E or K at PB2 position 627 are active<sup>11,13</sup>. Influenza A viruses that circulate in humans either have lysine at PB2 627 or, as in 2009 pandemic H1N1 virus, substitutions at residues 271, 590 and 591 (refs 14, 15) that map close to 627 on the polymerase structure<sup>16</sup>, suggesting a common mechanism of host adaptation.

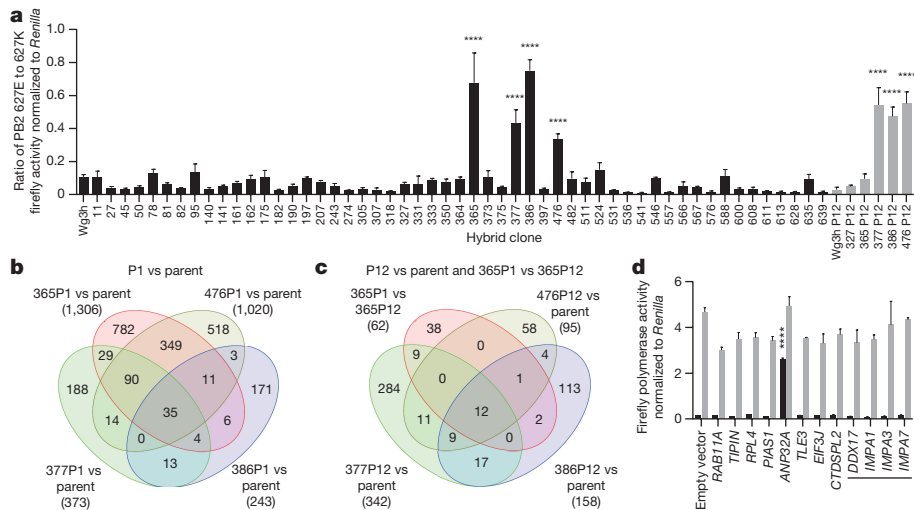
We showed previously that heterokaryons formed by fusing human and avian cells supported avian virus polymerase activity, suggesting that host range restriction was explained by a species-specific difference

in a positive avian host factor that was either lacking or different in mammals<sup>10</sup>. To identify this factor, we took advantage of a chicken genome radiation hybrid panel based on the Wg3H hamster cell line and harbouring different fragments of chicken chromosomes<sup>17</sup>. Each radiation hybrid clone was screened for the ability to support activity of a H5N1 avian influenza virus polymerase (A/turkey/England/50-92/91 virus, hereafter called 50-92) compared with its human-adapted isoform bearing PB2(E627K). Four out of 53 clones supported activity of the avian virus polymerase (Fig. 1a). Owing to instability of the chicken chromosome content, one positive clone lost this phenotype by passage 12. To identify chicken genes expressed in the positive radiation hybrid clones but not in negatives, total RNA was analysed using an Affymetrix chicken gene array. The four early passage positive clones had 35 chicken genes in common (Fig. 1b). After P12, the remaining positive hybrids shared only 12 genes in common, all present on a region of chicken chromosome 10 (Fig. 1c and Extended Data Table 1). Genetic instability of the clones is revealed by principal component analysis (PCA) mapping analysis, as is variation between negative clones (Extended Data Fig. 1).

We targeted candidate genes within this region and co-expressed each individually in human 293T cells with reconstituted H5N1 50-92 influenza virus polymerase. Expression of a single chicken gene, *chANP32A* (chicken acidic (leucine-rich) nuclear phosphoprotein 32 family, member A), rescued avian virus polymerase activity to levels similar to human-adapted PB2 627K viral polymerase (Fig. 1d). All positive hybrid clones expressed *chANP32A* (Extended Data Fig. 2). Knockdown of *chANP32A* abrogated the ability of positive radiation hybrid clone 476 to support avian virus polymerase, suggesting that this single chicken gene was responsible for the phenotype (Extended Data Fig. 3). Chicken homologues of DDX17 and members of the importin alpha family previously implicated in host range restriction of influenza virus polymerase did not rescue avian virus polymerase activity in human cells (Fig. 1d)<sup>18,19</sup>.

ANP32A is a member of a family of nuclear proteins implicated in multiple cellular pathways, including transcriptional regulation by chromatin remodelling, messenger RNA export and cell death<sup>20</sup>. ANP32 proteins share conserved amino-terminal domains, comprised of leucine rich repeats (LRR) and C-terminal low complexity acidic regions (LCAR) comprised of 60–75% glutamic or aspartic acid (Fig. 4a and Extended Data Fig. 8). We expressed cloned human homologue (*huANP32A*) as well as human and chicken genes for the related family member ANP32B. Expression of *chANP32A*, but not *huANP32A*, in human cells rescued activity of several different avian influenza polymerases including those from low (H1N1 A/duck/Bavaria/79 (Bav), H9N2 A/UDL/2008 (UDL)), and high (H5N1 (A/turkey/Turkey/05/2005 (Ty05) and A/turkey/England/50-92/91 (50-92)) pathogenicity avian viruses, and a human H3N2 virus (A/Victoria/3/75 (Vic)) when PB2 bore 627E (Extended Data Fig. 4). Expression of

<sup>1</sup>Section of Virology, Department of Medicine, Imperial College London, St Mary's Campus, London W2 1PG, UK. <sup>2</sup>Centre d'études d'agents Pathogènes et Biotechnologies pour la Santé (CPBS), FRE 3689, CNRS-UM, 34293 Montpellier, France. <sup>3</sup>Avian Viral Diseases Programme, The Pirbright Institute, Ash Road, Pirbright, Woking GU24 0NF, UK. <sup>4</sup>UMR INRA/Génétique Physiologie et Systèmes d'Élevage, INRA, 31326 Castanet-Tolosan, France.



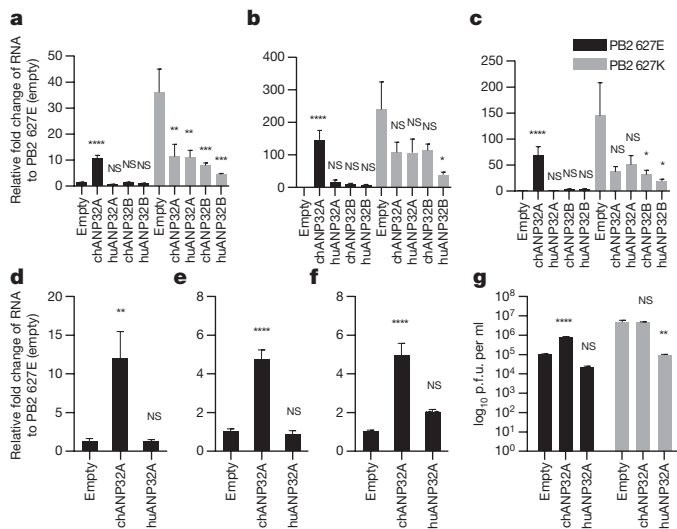
**Figure 1 | Identification of a positive avian cellular factor that permits avian influenza virus polymerase activity in mammalian cells by screening radiation hybrid clones.** **a**, Radiation hybrid (RH) clones transfected with mouse-poll-firefly minigenome reporter, avian influenza virus H5N1 50-92 polymerase (either PB2 627E or PB2 627K) and *Renilla* expression control. Passage 1 clones black bars, passage 12 clones grey bars. (Data as a ratio of PB2 627E over PB2 627K polymerase activity (firefly normalized to *Renilla*);  $n = 3$  biological replicates; error as s.e.m. of the ratio; one-way ANOVA, comparisons to Wg3h, \*\*\*\* $P < 0.0001$ ). **b**, Venn diagram of microarray data analysis of P1 positive hybrid clones vs parent Wg3h cells. **c**, Venn diagram of P12 positive hybrid clones vs parent and 365P1 (positive RH clone) vs 365P12 (reverted RH clone) (total gene numbers in brackets); Two-way ANOVA (variables: clone

and passage number) adjusted by Benjamini–Hochberg multiple-testing correction (false discovery rate (FDR) of  $P < 0.05$ ). Statistically significant genes identified and those with fold-change values  $\leq \pm 1.5$  removed. **d**, Cloned chicken genes from chromosome 10 or chicken homologues of genes previously implicated with PB2 host range (underlined) expressed in 293T cells for 20 h before transfection with pHOM1-firefly minigenome reporter, avian virus 50–92 polymerase (either PB2 627E (Black bars) or PB2 627K (grey bars)) and *Renilla* expression control. Luciferase activity was assayed after a further 24 h. (Data are firefly activity normalized to *Renilla*;  $n = 3$  biological replicates; error as s.e.m.; one-way ANOVA, comparisons to empty vector, \*\*\*\* $P < 0.0001$ , pattern of results consistent in at least three independent experiments.).

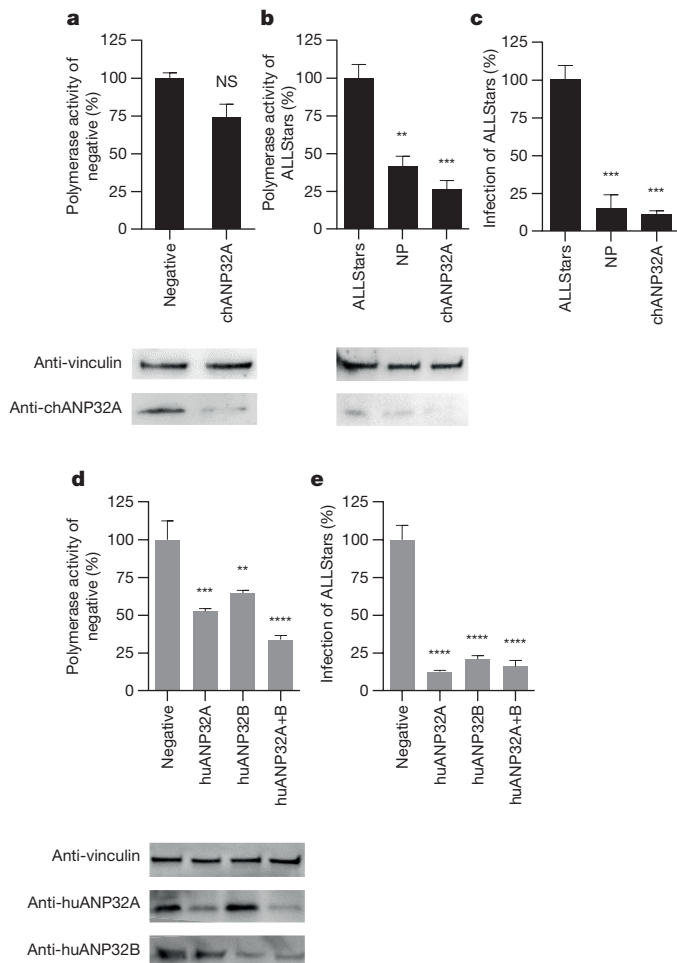
chANP32A, but not huANP32A, chANP32B or huANP32B, increased levels of all three species of minigenome RNA produced by reconstituted avian viral polymerase in human cells (Fig. 2a–c). Increase in levels of viral genomic RNA (vRNA) and complementary antigenomic RNA (cRNA) in the polymerase assay indicates that RNA replication, and not only mRNA transcription or stability, was enhanced by the expression of the chicken host factor in mammalian cells. Expression of chANP32A in human cells did not affect levels or nuclear accumulation of PB2, suggesting that its ability to support polymerase activity in the cell nucleus is not mediated by enhanced nuclear trafficking (Extended Data Fig. 5).

Additionally, expression of chANP32A enhanced avian virus replication in human cells. Levels of all three influenza RNA species derived from avian virus segment 1 were increased by expressed chANP32A but not by huANP32A (Fig. 2d–f). Moreover, the yield of infectious avian-like influenza virus was enhanced whether chANP32A was expressed in human cells at 33 °C (Fig. 2g) or 37 °C (Extended Data Fig. 4c). In contrast, chANP32A did not enhance yields of an isogenic PB2 627K recombinant virus, and overexpression of huANP32A reduced its titre (Fig. 2g and Extended Data Fig. 4c).

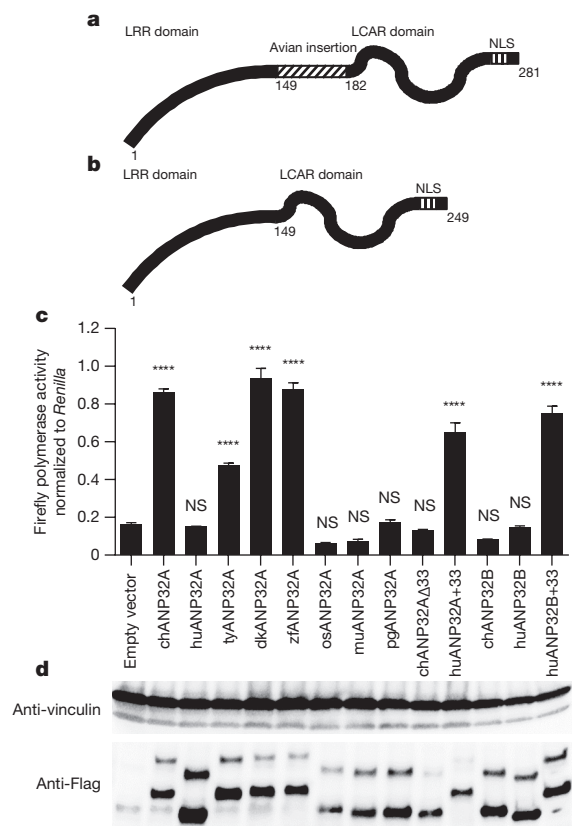
We confirmed a role for ANP32A in supporting virus replication in avian cells. In chicken cells (DF-1, immortalized chicken fibroblasts),



**Figure 2 | Expression of chANP32A in human cells rescues transcription and replication of avian influenza virus.** **a–c**, ch or huANP32A, ANP32B or empty vector were expressed for 20 h before transfection of avian virus 50–92 polymerase (either PB2 627E (black bars) or PB2 627K (grey bars)), and incubated at 37 °C for 24 h before reverse transcription followed by quantitative PCR (qRT–PCR) for luciferase gene of the viral minigenome reporter, vRNA (**a**), mRNA (**b**) and cRNA (**c**). **d–f**, 293T cells expressing chANP32A, huANP32A or empty vector for 20 h before infection by avian influenza virus (H9N2 A/chicken/UDL-01/2008) virus (multiplicity of infection (MOI) 1.0), incubated at 37 °C for 24 h before qRT–PCR for viral segment 1; vRNA (**d**), mRNA (**e**) and cRNA (**f**). (Data expressed as fold change to empty vector, normalized to 18S RNA, calculated by  $\Delta\Delta C_T$ ;  $n = 3$  biological replicates; error plotted as s.e.m.; one-way ANOVA, comparisons to empty vector; NS = not significant, \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$ ; pattern of results consistent in at least three independent experiments.). **g**, 293T cells expressing ch or huANP32A or empty vector for 20 h before infection (MOI 0.1) with avian-like influenza virus (recombinant PR8 virus with H5N1 Ty05 polymerase, M and NS gene segments and PR8 HA and NA genes) bearing PB2 627E (black bars) or PB2 627K (grey bars). Cells incubated at 33 °C and cell supernatant titrated for infectious virus 24 h post infection on MDCK cells by plaque assay. (Data displayed as  $\log_{10}$ (plaque forming units per ml);  $n = 3$  biological replicates; error plotted as s.e.m.; one-way ANOVA, comparisons to empty vector; NS, not significant, \*\* $P < 0.01$ , \*\*\*\* $P < 0.0001$ ; pattern of results consistent in at least three independent experiments.).



**Figure 3 | Knockdown of ANP32 reveals avian influenza polymerase dependence on chANP32A and dependence on huANP32A and B in human cells by human-adapted influenza polymerase.** **a**, DF-1 cells transduced with vesicular stomatitis virus (VSV)-G lentiviral vectors delivering transgenes expressing puromycin and shRNA targeting chANP32A or negative. Puromycin-selected cells were transfected with pCOM1-firefly minigenome reporter, avian 50-92 polymerase (627E) and *Renilla* expression control. **b**, siRNA (100 nM) applied to DF-1 chANP32A shRNA cells. After 48 h, cells transfected with avian 50-92 polymerase (627E), minigenome reporter and *Renilla* expression control. Luciferase activity measured 20 h later. Knockdown in chicken DF-1 cells verified by immunoblotting using antibody against vinculin and chANP32A. **c**, DF-1 cells depleted of chANP32A by siRNA infected with avian-like influenza virus (PR8 virus bearing H5N1 Ty05 polymerase genes with PB2 627E, MOI 0.01). 24 h later cell supernatants were titrated for infectious virus by plaque assay on MDCK cells. **d**, 293T cells transduced with lentiviral vectors delivering transgenes expressing puromycin and shRNA targeting huANP32A, huANP32B, both huANP32A and B, or negative. Puromycin-selected cells were transfected with pCOM1-firefly minigenome reporter, human-adapted avian 50-92 polymerase (627K) and *Renilla* expression control. Luciferase activity measured after 20 h. Knockdown in 293T cells was verified by immunoblotting using antibody against vinculin, huANP32A and huANP32B. (**a**, **b**, **d**, Data are firefly activity normalized to *Renilla*, plotted as per cent of negative or ALLStars;  $n = 3$  biological replicates; error as s.e.m.; one-way ANOVA comparisons to ALLStars or negative; NS= not significant,  $**P < 0.01$ ,  $***P < 0.001$ ,  $****P < 0.0001$ ). **e**, Puromycin-selected A549 cells expressing shRNA against huANP32A and/or B were infected with human (H3N2) virus A/England/691/2010 (MOI 0.1). After 24 h, cell supernatants were titrated by plaque assay on MDCK cells. (**c**, **e**, Data are % p.f.u. relative to ALLStars or negative;  $n = 3$  biological replicates; error as s.e.m.; one-way ANOVA, comparisons to ALLStars or negative;  $***P < 0.001$ ,  $****P < 0.0001$ ; pattern of results consistent in at least three independent experiments.).



**Figure 4 | Activity of avian influenza virus polymerase is dependent on a unique amino acid sequence present on avian ANP32A proteins.** **a**, chANP32A protein schematic with 33 amino acid insertion relative to human homologue. **b**, huANP32A schematic representative of ANP32A from mammals, ostrich and of ANP32B. **c**, 293T cells transfected with Flag-tagged ANP32 and after 20 h transfected with pCOM1-firefly minigenome reporter, avian 50-92 polymerase (627E) and *Renilla* expression control. Luciferase activity measured 24 h later. (Data are PB2 627E polymerase activity normalized to *Renilla*;  $n = 3$  biological replicates; error plotted as s.e.m. of the ratio; one-way ANOVA, comparisons to empty vector; NS= not significant,  $****P < 0.0001$ ; pattern of results consistent in at least three independent experiments.) ty, turkey; dk, duck; zf, zebra finch; os, ostrich; mu, mouse; pg, pig. **d**, Immunoblot analysis of Flag-tagged ANP32A constructs using antibody against Flag peptide and vinculin.

short hairpin RNA (Fig. 3a and Extended Data Fig. 6) and short interfering RNA (Fig. 3b) targeting chANP32A decreased activity of avian virus polymerase (PB2 627E). Knockdown of ANP32A in DF-1 cells also decreased the yield of infectious avian-like influenza virus (Fig. 3c).

We next investigated whether the human homologues of ANP32A or B support activity of influenza polymerase in human cells. Human ANP32A and B were previously listed in the human influenza A virus interactome<sup>21</sup>, and depletion of huANP32A and/or huANP32B from human cells decreased polymerase activity and replication of human influenza virus<sup>22,23</sup>. We used sh or siRNA to target huANP32A and B in human cells. Human-adapted viral polymerase (50-92 PB2 627K) activity was reduced when levels of either ANP32A or B were decreased, and further reduced in cells from which both proteins were depleted, suggesting a dependence on both family members (Fig. 3d and Extended Data Fig. 7). Yields of influenza virus A/England/691/2010, a seasonal H3N2 human virus, were decreased in cells depleted of ANP32A and/or ANP32B (Fig. 3e). We note with interest that both depletion and overexpression of human ANP32A or B were deleterious for polymerase activity and virus yields in human cells (Fig. 2), which suggests that influenza virus relies on balanced expression of ANP32 proteins.

Sequence alignment of avian and mammalian ANP32A revealed that avian (except ratite) ANP32A harbours an additional 33 amino



acids (176–208), derived partly by a repeat of amino acids 149–175, expressed from an additional exon absent from mammals and ostrich (Extended Data Fig. 8). Avian ANP32B does not contain the additional sequence (Extended Data Fig. 8). We cloned and Flag-tagged ANP32A cDNA from chicken, turkey, duck, and zebra finch, all of which encode homologues with the additional sequence, and also from ostrich, human, mouse, and pig, all of which were mammalian-like in length (Fig. 4a, b, d). The four longer avian proteins increased avian, but not human-adapted, influenza virus polymerase activity in human cells (Fig. 4c and Extended Data Fig. 9). In contrast, the ostrich protein, or any of the mammalian homologues, did not rescue avian virus polymerase activity. Interestingly, replication of avian influenza viruses in ostrich and other raptorial selects for mammalian-adapted PB2 mutants<sup>24</sup>.

Deletion of amino acids 176–208 from chANP32A abrogated its ability to support avian virus polymerase. Conversely, insertion of the additional avian-specific 33 amino acids bestowed upon either huANP32A or huANP32B the ability to support avian virus polymerase (Fig. 4c, d).

In summary, we conducted a novel screen of chicken genome radiation hybrid cells and identified a single chicken gene, *ANP32A*, that supports avian influenza virus polymerase activity in human cells. We suggest that avian influenza virus has co-evolved with its natural hosts, wild birds, to co-opt avian ANP32A as a host factor that supports its polymerase activity in the nucleus. The avian influenza polymerase cannot efficiently utilize shorter ANP32 proteins, such as ANP32B or mammalian ANP32A homologues, for this activity. Acquisition of host-adapting mutations, such as PB2(E627K), enables polymerase activity to be supported by the shorter ANP32 proteins typical of mammalian hosts. The underlying mechanism of avian influenza polymerase restriction in human cells has evaded researchers for decades. Others have previously demonstrated a role for PB2 residue 627 in its interaction with NP<sup>12,25</sup>, and viral RNA promoter sequences<sup>26,27</sup> that may influence recognition of incoming vRNPs by RIG-I<sup>28</sup>. The nature of residue 627 has also been linked with a PB2 dependency on specific importin family members<sup>19,29</sup>. How these observations are linked with the species difference in host factor ANP32A we describe here remains to be elucidated.

Substitutions in PB2 are rapidly selected after avian influenza viruses enter mammalian hosts and PB2 adaptation is essential to support the pandemic potential of emerging avian influenza viruses. This stringent requirement for the virus to optimize replication efficiency in the human host suggests that disrupting the interaction of the virus with ANP32A may be a novel means for virus control.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 22 June; accepted 23 November 2015.**

- Neumann, G. & Kawaoka, Y. Transmission of influenza A viruses. *Virology* **479–480**, 234–246 (2015).
- Cauldwell, A. V., Long, J. S., Moncorgé, O. & Barclay, W. S. Viral determinants of influenza A virus host range. *J. Gen. Virol.* **95**, 1193–1210 (2014).
- Almond, J. W. A single gene determines the host range of influenza virus. *Nature* **270**, 617–618 (1977).
- Subbarao, E. K., London, W. & Murphy, B. R. A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. *J. Virol.* **67**, 1761–1764 (1993).
- Naffakh, N., Tomoiu, A., Rameix-Welti, M.-A. & van der Werf, S. Host restriction of avian influenza viruses at the level of the ribonucleoproteins. *Annu. Rev. Microbiol.* **62**, 403–424 (2008).
- Mänz, B., Schwemmle, M. & Brunotte, L. Adaptation of avian influenza A virus polymerase in mammals to overcome the host species barrier. *J. Virol.* **87**, 7200–7209 (2013).
- Eisfeld, A. J., Neumann, G. & Kawaoka, Y. At the centre: influenza A virus ribonucleoproteins. *Nature Rev. Microbiol.* **13**, 28–41 (2015).
- Steel, J., Lowen, A. C., Mubareka, S. & Palese, P. Transmission of influenza virus in a mammalian host is increased by PB2 amino acids 627K or 627E/701N. *PLoS Pathog.* **5**, e1000252 (2009).
- Van Hoeben, N. et al. Human HA and polymerase subunit PB2 proteins confer transmission of an avian influenza virus through the air. *Proc. Natl Acad. Sci. USA* **106**, 3366–3371 (2009).

- Moncorgé, O., Mura, M. & Barclay, W. S. Evidence for avian and human host cell factors that affect the activity of influenza virus polymerase. *J. Virol.* **84**, 9978–9986 (2010).
- Long, J. S. et al. The effect of the PB2 mutation 627K on highly pathogenic H5N1 avian influenza virus is dependent on the virus lineage. *J. Virol.* **87**, 9983–9996 (2013).
- Mehle, A. & Doudna, J. A. An inhibitory activity in human cells restricts the function of an avian-like influenza virus polymerase. *Cell Host Microbe* **4**, 111–122 (2008).
- Massin, P., van der Werf, S. & Naffakh, N. Residue 627 of PB2 is a determinant of cold sensitivity in RNA replication of avian influenza viruses. *J. Virol.* **75**, 5398–5404 (2001).
- Bussey, K. A., Bousse, T. L., Desmet, E. A., Kim, B. & Takimoto, T. PB2 residue 271 plays a key role in enhanced polymerase activity of influenza A viruses in mammalian host cells. *J. Virol.* **84**, 4395–4406 (2010).
- Mehle, A. & Doudna, J. A. Adaptive strategies of the influenza virus polymerase for replication in humans. *Proc. Natl Acad. Sci. USA* **106**, 21312–21316 (2009).
- Pflug, A., Guilligay, D., Reich, S. & Cusack, S. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* **516**, 355–360 (2014).
- Morisson, M. et al. ChickRH6: a chicken whole-genome radiation hybrid panel. *Genet. Sel. Evol.* **34**, 521–533 (2002).
- Bortz, E. et al. Host- and strain-specific regulation of influenza virus polymerase activity by interacting cellular proteins. *MBio* **2**, e00151–11 (2011).
- Gabriel, G. et al. Differential use of importin- $\alpha$  isoforms governs cell tropism and host adaptation of influenza virus. *Nature Commun.* **2**, 156 (2011).
- Reilly, P. T., Yu, Y., Hamiche, A. & Wang, L. Cracking the ANP32 whips: important functions, unequal requirement, and hints at disease implications. *BioEssays* **36**, 1062–1071 (2014).
- Bradel-Trethaway, B. G. et al. Comprehensive proteomic analysis of influenza virus polymerase complex reveals a novel association with mitochondrial proteins and RNA polymerase accessory factors. *J. Virol.* **85**, 8569–8581 (2011).
- Watanabe, T. et al. Influenza virus-host interactome screen as a platform for antiviral drug development. *Cell Host Microbe* **16**, 795–805 (2014).
- Sugiyama, K., Kawaguchi, A., Okuwaki, M. & Nagata, K. pp32 and APRIL are host cell-derived regulators of influenza virus RNA synthesis from cRNA. *eLife* **4**, e08939 (2015).
- Shinya, K. et al. Ostrich involvement in the selection of H5N1 influenza virus possessing mammalian-type amino acids in the PB2 protein. *J. Virol.* **83**, 13015–13018 (2009).
- Rameix-Welti, M.-A., Tomoiu, A., Dos Santos Afonso, E., van der Werf, S. & Naffakh, N. Avian Influenza A virus polymerase association with nucleoprotein, but not polymerase assembly, is impaired in human cells during the course of infection. *J. Virol.* **83**, 1320–1331 (2009).
- Crescenzo-Chaigne, B., van der Werf, S. & Naffakh, N. Differential effect of nucleotide substitutions in the 3' arm of the influenza A virus vRNA promoter on transcription/replication by avian and human polymerase complexes is related to the nature of PB2 amino acid 627. *Virology* **303**, 240–252 (2002).
- Paterson, D., de Velthuis, A. J. W., Vreede, F. T. & Fodor, E. Host restriction of influenza virus polymerase activity by PB2 627E is diminished on short viral templates in a nucleoprotein-independent manner. *J. Virol.* **88**, 339–344 (2014).
- Weber, M. et al. Influenza virus adaptation PB2-627K modulates nucleocapsid inhibition by the pathogen sensor RIG-I. *Cell Host Microbe* **17**, 309–319 (2015).
- Resa-Infante, P. et al. The host-dependent interaction of  $\alpha$ -importins with influenza PB2 polymerase subunit is required for virus RNA replication. *PLoS One* **3**, e3904 (2008).

**Acknowledgements** We thank G. Maertens, J. Stech, R. Fouchier, A. Cauldwell, G. Roche, J. McCauley, D. Huntley, A. Vaughan, V. Nair and H. Shelton for provision of reagents, advice and discussions. This work was funded by BBSRC sLoLa BB/K002465/1 “Developing Rapid Responses to Emerging Virus Infections of Poultry (DDREVIP)” which funds J.S.L. and E.S.G., B.M. was funded by a Wellcome Trust studentship. R.F. and O.M. were funded by a Wellcome Trust Programme Grant (087039/Z/08/Z). O.M. was funded by MRC (G0600006). M.I. was funded by a BBSRC Avian Diseases Programme Grant (BBS/E/1/00001708).

**Author Contributions** J.S.L. designed and performed the experiments and wrote the manuscript. E.S.G. performed microarrays and analysed data. O.M. generated plasmids for polymerase assays and wrote the manuscript. R.F., E.S.G. and B.M. performed qRT-PCR analysis. J.J. generated plasmids for polymerase assays. M.I. supplied UDL/08 reverse genetics system. A.V. and M.M. supplied radiation hybrid clones. M.A.S. analysed data, designed microarray experiments and wrote the manuscript. W.S.B. designed experiments and wrote the manuscript.

**Author Information** The microarray data have been submitted to The European Bioinformatics Institute (EBI) (<http://www.ebi.ac.uk>) ArrayExpress under accession number E-MTAB-3643. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.S.B. ([w.barclay@imperial.ac.uk](mailto:w.barclay@imperial.ac.uk)).

## METHODS

**Cells and cell culture.** Chinese Hamster cell line (Wg3h), chicken genome RH clones (Laboratoire de génétique cellulaire, Institut National de la Recherche Agronomique, Castanet-Tolosan, France)<sup>17</sup>, human embryonic kidney (293T) (ATCC), human lung adenocarcinoma epithelial cells (A549) (ATCC) and Madin-Darby canine kidney (MDCK) cells (ATCC) were maintained in cell culture media (Dulbecco's modified Eagle's medium (DMEM; Invitrogen) supplemented with 10% fetal calf serum (FCS) (Biosera) and with 1% penicillin-streptomycin (Invitrogen)). Chicken fibroblast (DF-1) (ATCC) cells were maintained in DF-1 cell culture media (DMEM supplemented with 10% FCS, 5% tryptose phosphate broth (Sigma-Aldrich) and 0.1% penicillin-streptomycin (Invitrogen)). Cell lines were maintained at 37°C in a 5% CO<sub>2</sub> atmosphere. Cell lines were authenticated by RT-PCR and verified negative for mycoplasma.

**Microarray analysis.** Total RNA was extracted from  $1 \times 10^6$  cells that were of the same passage as those tested for polymerase, using an RNeasy kit (Qiagen) according to the manufacturer's instructions. On-column DNA digestion was performed using RNase-free DNase (Qiagen) to remove contaminating genomic DNA. RNA samples were quantified using a Nanodrop Spectrophotometer (Thermo Scientific) and checked for quality using a 2100 Bioanalyzer (Agilent Technologies). All RNA samples had an RNA integrity number (RIN)  $\geq 9.8$ . Array hybridization was performed according to the manufacturer's instructions (Affymetrix). Labelled samples were hybridized to the Affymetrix Chicken Gene 1.0 ST Arrays in a GeneChip Hybridization Oven for 16 h at 45°C and 60 rpm in an Affymetrix Hybridization Oven 645. After washing and staining, the arrays were scanned with the Affymetrix GeneChip Scanner 3000 7G. Gene-level expression signal estimates were derived from CEL files generated from raw data using the multi-array analysis (RMA) algorithm implemented from the Affymetrix GeneChip Command Console Software Version 3.0.1. Data Pre-Processing and filtering was done using Partek Software Version 6.6 and included: RMA background correction, quantile normalization across all chips in the experiment, log<sub>2</sub> transformation and median polish summarization. Statistically significant genes were discovered by comparisons between the positive RH clones, parent cells and negative RH clones by two-way ANOVA (variables: clone and passage number) adjusted with the Benjamini-Hochberg multiple-testing correction (false discovery rate (FDR) of  $P < 0.05$ ). Statistically significant genes were identified and those with fold-change values  $< |\pm 1.5|$  were removed.

**Cloning of candidate cDNAs.** Sequence specific primers were used to amplify targeted cellular transcripts of chicken genes from chromosome 10, including: ANP32A, RAB11A (Ras-related protein Rab-11A), TIPIN (TIMELESS Interacting Protein), RPL4 (Ribosomal Protein L4), PIAS1 (Protein Inhibitor Of Activated STAT, 1), TLE3 (Transducin-Like Enhancer Of Split 3), EIF3J (Eukaryotic Translation Initiation Factor 3, Subunit J) and CTDSP2 (CTD (Carboxy-Terminal Domain, RNA Polymerase II, Polypeptide A) Small Phosphatase Like 2) from total RNA extracted from RH clone 476, using SuperScript III One-Step RT-PCR System (Invitrogen). Chicken ANP32B, DDX17 (DEAD (Asp-Glu-Ala-Asp) Box Helicase 17), IMP $\alpha$ 1, 3 and 7 (Importin  $\alpha$ -1,3 and 7) cDNAs were amplified from RNA extracted from DF-1 cells. PCR products were cloned into the pCAGGS expression vector. cDNAs of full length ANP32A isoforms of several species were generated by gene synthesis (GeneArt Strings DNA Fragments) and inserted into the pCAGGS expression vector, based on the following sequences: Chicken ANP32A (chANP32A) (*Gallus gallus*, XP\_413932.3), human ANP32A (huANP32A) (*Homo sapiens*, NP\_006296.1), zebra finch ANP32A (zfANP32A) (*Taeniopygia guttata*, XP\_012424064.1), duck ANP32A (dkANP32A) (*Anas platyrhynchos*, XP\_005023024.1), turkey ANP32A (tyANP32A) (*Meleagris gallopavo*, XP\_010715918.1), ostrich ANP32A (osANP32A) (*Struthio camelus australis*, XP\_009665579.1), pig ANP32A (pgANP32A) (*Sus scrofa*, XP\_003121807.3), mouse ANP32A (muANP32A) (*Mus musculus*, NP\_033802.2), chicken ANP32B (chANP32B) (*Gallus gallus*, NP\_001026105.1) and human ANP32B (huANP32B) (*Homo sapiens*, NP\_006392.1). The sequence of dkANP32A was amended to contain the sequence encoding the intact N terminus by comparison with duck RNA-seq data (from ENA\_ERP005909) that had been *de novo* assembled (using CLC Genomic Workbench 7.5.1), where the appropriate contig was identified by BLASTX against the chicken ANP32A protein sequence. Furthermore, the duck sequence was confirmed by reverse-transcription of ANP32A mRNA derived from duck embryonic fibroblast cells and DNA sequencing. Mutants of these sequences were also generated by gene synthesis, as above, and included: chANP32A $\Delta$ 33 (chANP32A with the 'avian insertion' deleted (aa176–208)), huANP32A+33 and huANP32B+33 (huANP32A and B with the avian sequence –VLSLVKDRDDKEAPDSDAEGYVEGLDDEEDED– inserted after aa175 (huANP32A) and aa173 (huANP32B)). All plasmid constructs were verified by DNA sequencing. Primers and sequence information are available upon request.

**Generation of recombinant and clinical Influenza A virus.** Reverse genetics systems for the following virus strains were used in this study: PR8 (H1N1 A/Puerto

Rico/8/1934)<sup>30</sup>, UDL (H9N2 A/chicken/UDL-01/08)<sup>31</sup> (developed in collaboration with M. Iqbal, Pirbright Institute, UK) and Ty05 (H5N1 A/Turkey/Turkey/5/2005) (a kind gift from R. Fouchier, Erasmus University, Netherlands). The PB2(E627K) substitution was made to 50-92 and Ty05 (K627E) by overlapping PCR of the PB2 plasmid as previously described<sup>11</sup>. Ty05:PR8 6:2 recombinant virus was generated with the HA and NA derived from PR8 and the internal genes of Ty05; virus rescue was performed by co-transfection of the 12-plasmid system: 8 polI plasmids as described above and 4 helper expression plasmids encoding A/Victoria/3/75 (VIC) polymerase components and NP expressed by the pCAGGS vector, as previously described<sup>32,33</sup>. UDL virus rescue was performed by co-transfection of a 9 plasmid system including: 7 bidirectional pHW2000 plasmids<sup>34</sup> encoding PB2, PA, HA, NA, NP, NS and M genes, together with a polI plasmid encoding PB1 and a pCAGGS expression plasmid of the UDL-PB1 gene. Ty05 virus stocks were propagated in 9-day-old embryonated chicken eggs incubated at 37°C. UDL virus stocks were generated in MDCK cells with infection media (serum free DMEM supplemented with 1% penicillin-streptomycin and  $1 \mu\text{g ml}^{-1}$  TPCK-treated trypsin (Lorne Labs)) and incubated at 37°C. Clinical isolate A/England/691/2010 (H3N2) (Public Health England) was propagated in MDCK cells with infection media. Aliquots of infectious virus were stored at  $-80^\circ\text{C}$ . Infectious titres were determined by plaque assay on MDCK cells.

**Influenza A virus infection.** 293T cells were transfected with ANP32A (500 ng, 24-well) using Lipofectamine 3000 (Invitrogen) and after 20 h infected with virus diluted in serum free DMEM for 1 h at 33 or 37°C (MOI as indicated in the relevant figure legends) and replaced with cell culture medium (for qRT-PCR analysis) or DMEM supplemented with 0.1% FCS and  $1 \mu\text{g ml}^{-1}$  TPCK trypsin (Worthington-Biochemical) (for infectious virus titres). shRNA A549 cells were infected with virus as for 293T cells, except infection media lacked FCS. Infected cell lysates and cell supernatants were harvested at 12 and/or 24 h post-infection. Infectious titres were determined by plaque assay on MDCK cells.

**Safety/biosecurity.** All work with infectious agents was conducted in biosafety level 2 facilities, approved by the Health and Safety Executive of the UK and in accordance with local rules, at Imperial College London, UK.

**Polymerase assay.** Influenza polymerase activity was measured by use of a minigenome reporter which contains the firefly luciferase gene flanked by the non-coding regions of the influenza NS gene segment, transcribed from a species-specific polI plasmid with a mouse terminator sequence<sup>35</sup>. The human and chicken polI minigenomes (pHOM1-Firefly and pCOM1-Firefly) are described previously<sup>36</sup>; pMouse-PolI-Firefly was generated by substituting in the mouse polI promoter sequence<sup>37</sup>. pCAGGS expression plasmids encoding each polymerase component and NP for 50–92 (H5N1 A/Turkey/England/50–92/91), Ty05, VIC and BAV (A/Duck/Bavaria/1/77) are described previously<sup>11,36,38</sup>. UDL PB1, PB2, PA and NP genes were sub-cloned into the pCAGGS plasmid. Mutagenesis of PB2 genes to encode PB2 627K or 627E was performed by overlapping PCR as described previously<sup>11,36</sup>. All plasmid constructs were verified by DNA sequencing. Primers and sequence information are available upon request. To measure influenza polymerase activity, 293T cells were transfected in 48-well plates with pCAGGS plasmids encoding the PB1 (20 ng), PB2 (20 ng), PA (10 ng) and NP (40 ng) proteins, together with 20 ng species-specific minigenome reporter and 10 ng *Renilla* luciferase expression plasmid (pCAGGS-*Renilla*)<sup>39</sup> as an internal control, using Lipofectamine 3000 transfection reagent (Invitrogen) according to manufacturers' instructions. Wg3h, RH clones and DF-1 cells were transfected as 293T cells but with twice the concentration of DNA and using Lipofectamine 2000 (Invitrogen). Cells were incubated at 37°C. 20 h after transfection, cells were lysed with 50  $\mu\text{l}$  of passive lysis buffer (Promega), and firefly and *Renilla* luciferase bioluminescence was measured using a Dual-luciferase system (Promega) with a FLUOstar Omega plate reader (BMG Labtech). The effect of cellular factors on influenza polymerase was examined by polymerase assay after expression of constructs (250 ng) for 24 h.

**shRNA mediated silencing.** Silencing was achieved by lentiviral delivery of shRNA encoding transgenes. Lentiviral vectors were generated using the TRC1.5-pLKO.1-puro plasmid (MISSION Sigma-Aldrich) containing the shRNA sequence and puromycin selection gene. shRNA sequences for target genes were as follows: huANP32A (TRCN0000006905, 5'-CCGGCCTGAAGATGAGGGAGA AGATCTCAGAGTCTTCTCCCTCATCTCAGGTTT-3' (target sequence, 5'-CCTGAAGATGAGGGAGAAGAT-3')), huANP32B (TRCN0000077928, 5'-CCGGCCACCCAAAGAGCCAAAGAATCTCGAGATTCTTTGGCTCTTTG GGTGGTTT-3' (target sequence, 5'-CCACCCAAAGAGCCAAAGAAT-3')), chANP32A (TRCN0000006902, 5'-CCGGCCTATTGTGATTGACTGTT TCTCGAGAAACAGTCAATCACAATGAGTTT-3' (target sequence, 5'-CCTATTGTGATTGACTGTTT-3')) and Negative (SHC002, Non-Mammalian shRNA Control) (MISSION Sigma-Aldrich). Lentiviruses were generated by co-transfection in 293T cells with pCMV-delta8.2<sup>40</sup>, pCAGG-VSV G<sup>41</sup> and TRC1.5-pLKO.1-puro at a ratio of 1:0.25:1 using Lipofectamine 3000 (Invitrogen), cell culture media was replaced after 16 hrs at a reduced volume and supernatant



harvested at 36 h post-transfection before being filtered (0.45 µm) and aliquots frozen at -80 °C. 293T, A549 or DF-1 cells were transduced with lentiviral vectors for 16 h before media was replaced. After 72 h incubation, cells were split and cell culture media was replaced containing 0.5 µg ml<sup>-1</sup> puromycin (Invivogen). Cells were incubated a further 72 h after selection before analysis.

**siRNA mediated silencing.** DF-1, 293T or RH clone 476 cells were transfected with 100 nM of siRNA using HiPerFect transfection reagent in 48-well plates, according to manufacturer's instructions (Qiagen). 48 h later cells were transfected with polymerase and minigenome constructs and harvested after a further 20 h, for luciferase quantification and knockdown analysis. Total RNA was extracted as described previously but with 100 µl of cell lysate added to AVL buffer before continuing with the RNeasy mini kit (Qiagen). siRNAs for target genes were as follows: AllStars Negative Control, huANP32A (SI02655212 FlexiTube), huANP32B (SI02655380 FlexiTube) (Qiagen), 50-92 NP (5'-AAGGAUCUUAUUCUUCGAG-3'), chANP32A (5'-GAGCTGGAATCTTGTAGTACA-3') (custom RNA oligos, Sigma-Aldrich).

**Quantification of chANP32A and B mRNA levels.** Total RNA from RH clones and DF-1 cells were extracted using an RNeasy mini kit (Qiagen), following manufacturer's instructions. During extraction of RNA, RNeasy columns were treated with RNase-Free DNase (Qiagen). RNA samples were quantified using a Nanodrop Spectrophotometer (Thermo Scientific). Equal concentrations of RNA were subject to first strand synthesis using QuantiTect Reverse Transcription Kit (Qiagen) with primers specific for chANP32A (5'-CAACTGTAGGTCATACGAAGGC-3') and chANP32B (5'-GGTGGCCTTGAAGTTCTAGC-3'). This product was then quantified with Mesa Green quantitative PCR (qPCR) MasterMix Plus for SYBR Assay I dTTP (Eurogentec) using the primers for first strand synthesis together with chANP32A (5'-GTTTGGCACTGAGGCTAAGC-3') and chANP32B (5'-ATGAGCATCGTCACCTCGC-3'). Real-time quantitative PCR analysis was performed (Applied Biosystems ViiA 7 Real-Time PCR System) and absolute copy numbers of either chANP32A or B calculated using a standard curve of known concentrations of the corresponding cDNA expression plasmid. Primers were designed to be specific to their target transcripts by using BLASTX against both the hamster and chicken genomes.

**Quantification of RNA generated by influenza polymerase.** Purified total RNA (1000 ng) was subject to first-strand cDNA synthesis with gene specific primers, oligo(dT)20 or random hexamers (to amplify mRNA) using SuperScript III (Invitrogen) followed by RNase H treatment (Invitrogen). Primer design was based on Obayashi *et al.* (2008)<sup>42</sup> for quantification of RNA species of the luciferase minigenome driven by reconstituted polymerase, and UDL PB2 RNA species were quantified using a tagged-primer system adapted from Kawakami *et al.* (2011)<sup>43</sup>. First strand primers included: luciferase vRNA (5'-TATGAACATTTTCGACGCTACCGTAGTGTT-3'), luciferase cRNA (5'-AGTAGAAACAAGGGTG-3'), luciferase mRNA (Oligo(dT)20), UDL PB2 vRNA (5'-GGCCGTCATGGTGGCGAAT<sub>(tag)</sub>GATGCGTGATGTATTGGGAAC-3'), UDL PB2 cRNA (5'-GCTAGCTTCAGCTAGGCATC<sub>(tag)</sub>AGTAGAAACAAGGTCGTT-3'), UDL PB2 mRNA (Oligo(dT)20) and 18S ribosomal RNA (Random Hexamers (Invitrogen)). After first strand synthesis, 1 µl of cDNA was subject to real-time quantitative PCR analysis with a gene specific primer pair using SYBR green PCR mix (Applied Biosystems) and analysed on the Applied Biosystems ViiA 7 Real-Time PCR System. Gene specific primer pairs were as follows: Luciferase gene (5'-CCGGAATGATTGTAGTGGCA-3' and 5'-TATGAACATTTTCGACGCTACCGTAGTGTT-3'), UDL PB2 vRNA (5'-GGCCGTCATGGTGGCGAAT<sub>(tag)</sub>-3' and 5'-CCTCTCAACACTGCAGATTCC-3'), UDL PB2 cRNA (5'-GCTAGCTTCAGCTAGGCATC<sub>(tag)</sub>-3' and 5'-GGAATCTGCAGTGTGAGAGG-3'), UDL PB2 mRNA (5'-GATGCGTGATGTATTGGGAAC-3' and 5'-CCTCTCAACACTGCAGATTCC-3') and 18S ribosomal RNA (5'-GCAAATTACCACTCCCG-3' and 5'-CTGCAGCACTTTAATATACGC-3'). Fold change RNA to PB2 627E with Empty vector was calculated by  $\Delta\Delta C_T$  including normalization to  $C_T$  values of 18S ribosomal RNA.

**Immunoblot analysis.** Cells were lysed in Passive Lysis buffer (Promega) or NP40 lysis buffer (for cellular fractionation) and prepared in Laemmli 2× buffer (Sigma-Aldrich). Cell proteins were resolved by SDS-PAGE using Mini-PROTEAN TGX Precast Gels (Bio-Rad). Immunoblotting was carried out using the following primary antibodies: anti-chANP32A rabbit polyclonal (LS-B10851, LifeSpan BioSciences, Inc.), anti-huANP32A rabbit polyclonal (AB51013, Abcam),

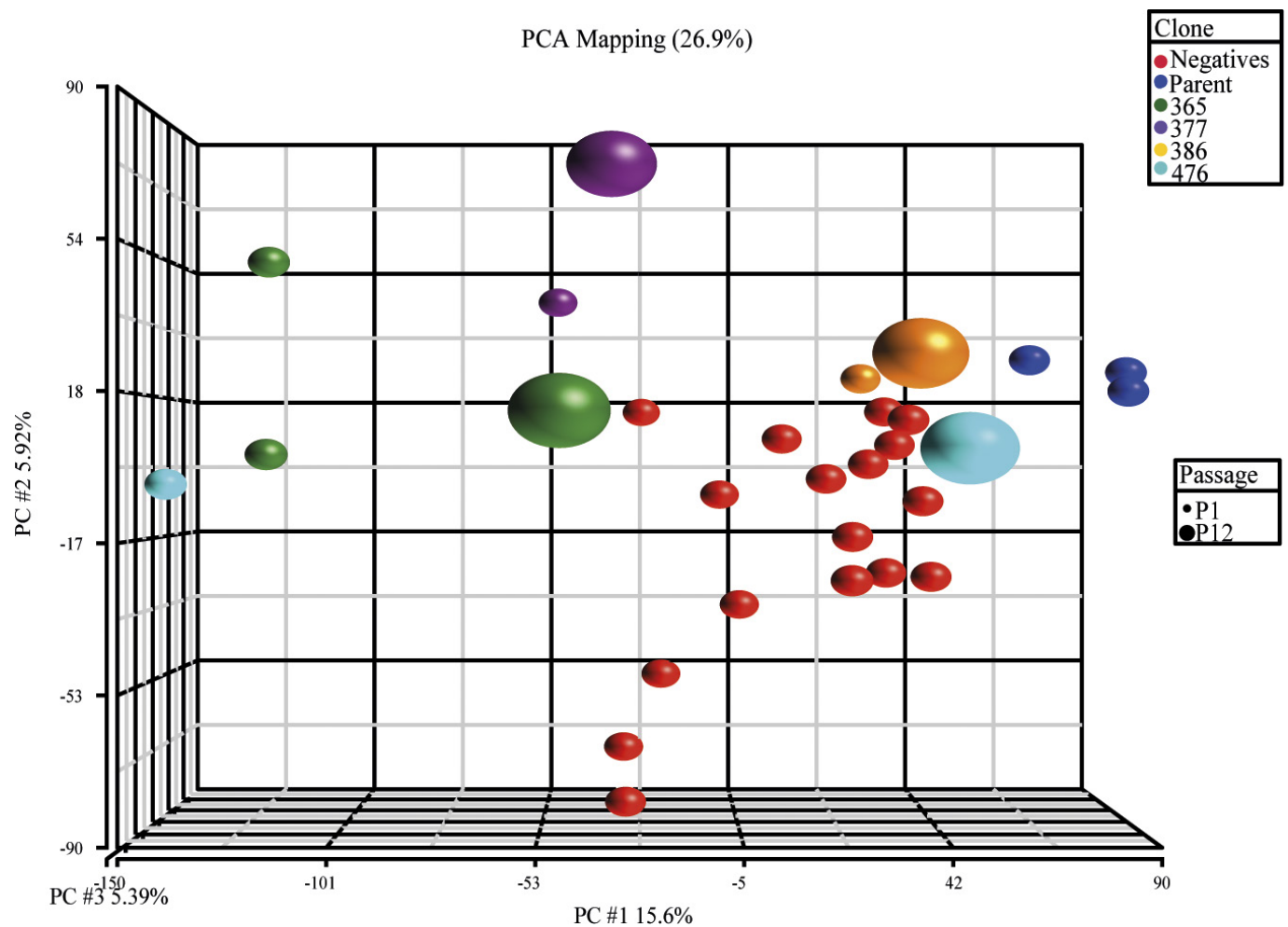
anti-huANP32B rabbit monoclonal (AB184565, Abcam),  $\alpha$ -vinculin rabbit monoclonal (AB129002, Abcam), anti-Flag M2 mouse monoclonal (F1804 or F3165, Sigma-Aldrich), anti-PB2 rabbit polyclonal (2N580, a kind gift from Paul Digard, Roslin Institute), and followed with secondary horseradish peroxidase-conjugated (HRP) antibodies: anti-mouse IgG (H/L):HRP goat polyclonal (STAR117P, AbD Serotec) and anti-rabbit IgG:HRP sheep polyclonal (STAR54, AbD Serotec). For quantification of cellular fractions, the following secondary antibodies were used: anti-rabbit IgG (H/L):DyLight 800 (5151P, Cell signalling) and anti-mouse IgG (H/L):DyLight 680 (5470P, Cell signalling). Protein bands were visualized by chemiluminescence (ECL+ western blotting substrate, Pierce) using a FUSION-FX imaging system (Vilber Lourmat).

**Cellular Fractionation.** 293T cells were transfected with empty vector or ANP32 plasmid together with the polymerase complex and NP of 50-92 (PB2 627E) and pHOM1-firefly minigenome reporter. After 24 h, cells monolayers were washed in ice-cold PBS and lysed in 0.1% NP40 buffer (50 mM Tris pH 7.5, 150 mM NaCl, 0.1% NP40 and protease inhibitors (EDTA-free COMPLETE tablet (Roche))). Total lysates were centrifuged at 228 g for 5 min at 4 °C. Supernatant was removed (Cytoplasmic fraction) and the nuclear pellet was resuspended in 1% NP40 (as above) and subject to syringing with a 25G needle. Fractions were analysed by immunoblotting.

**Bioinformatic and statistical analysis.** Analysis of Microarray data was performed as previously mentioned, using Affymetrix GeneChip Command Console Software Version 3.0.1. and Partek Software Version 6.6. Statistical analysis of biological replicates was performed by One-way ANOVA with Dunnett's multiple comparison analysis or Two-way ANOVA with Sidák multiple comparison analysis, using GraphPad Prism 6. Sequence alignments were performed using Geneious R6 software. Quantification of immunoblots was performed using Image Studio Lite V5.2.

No statistical methods were used to predetermine sample size. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

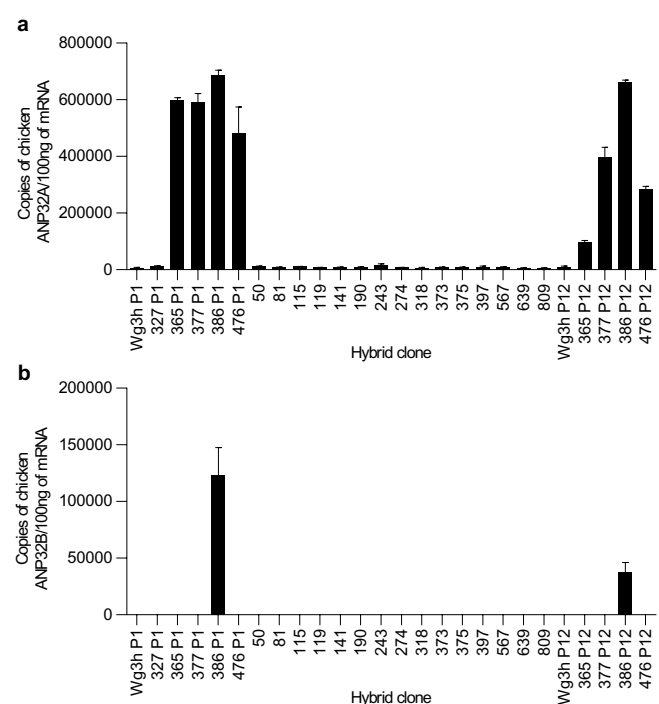
- Whiteley, A. *et al.* Generation of candidate human influenza vaccine strains in cell culture - rehearsing the European response to an H7N1 pandemic threat. *Influenza Other Respir. Viruses* **1**, 157-166 (2007).
- Iqbal, M., Yaqub, T., Mukhtar, N., Shabbir, M. Z. & McCauley, J. W. Infectivity and transmissibility of H9N2 avian influenza virus in chickens and wild terrestrial birds. *Vet. Res.* **44**, 100 (2013).
- Elleman, C. J. & Barclay, W. S. The M1 matrix protein controls the filamentous phenotype of influenza A virus. *Virology* **321**, 144-153 (2004).
- Neumann, G. *et al.* Generation of influenza A viruses entirely from cloned cDNAs. *Proc. Natl Acad. Sci. USA* **96**, 9345-9350 (1999).
- Hoffmann, E., Neumann, G., Kawaoka, Y., Hoborn, G. & Webster, R. G. A DNA transfection system for generation of influenza A virus from eight plasmids. *Proc. Natl Acad. Sci. USA* **97**, 6108-6113 (2000).
- Pleschka, S. *et al.* A plasmid-based reverse genetics system for influenza A virus. *J. Virol.* **70**, 4188-4192 (1996).
- Moncorgé, O. *et al.* Investigation of influenza virus polymerase activity in pig cells. *J. Virol.* **87**, 384-394 (2013).
- Flick, R. & Pettersson, R. F. Reverse genetics system for Uukuniemi virus (Bunyaviridae): RNA polymerase I-catalyzed expression of chimeric viral RNAs. *J. Virol.* **75**, 1643-1655 (2001).
- Howard, W. *et al.* Development of a reverse genetics system enabling the rescue of recombinant avian influenza virus A/Turkey/England/50-92/91 (H5N1). *Avian Dis.* **51** (Suppl. 1), 393-395 (2007).
- Cauldwell, A. V., Moncorgé, O. & Barclay, W. S. Unstable polymerase-nucleoprotein interaction is not responsible for avian influenza virus polymerase restriction in human cells. *J. Virol.* **87**, 1278-1284 (2013).
- Naldini, L., Blömer, U., Gage, F. H., Trono, D. & Verma, I. M. Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector. *Proc. Natl Acad. Sci. USA* **93**, 11382-11388 (1996).
- Ulm, J. W., Perron, M., Sodroski, J. & C Mulligan, R. Complex determinants within the Moloney murine leukemia virus capsid modulate susceptibility of the virus to Fv1 and Ref1-mediated restriction. *Virology* **363**, 245-255 (2007).
- Obayashi, E. *et al.* The structural basis for an essential subunit interaction in influenza virus RNA polymerase. *Nature* **454**, 1127-1131 (2008).
- Kawakami, E. *et al.* Strand-specific real-time RT-PCR for distinguishing influenza vRNA, cRNA, and mRNA. *J. Virol. Methods* **173**, 1-6 (2011).



**Extended Data Figure 1 | Analysis of mRNAs by PCA mapping reveals diversity of the radiation hybrid clones and their genetic instability during cell passage.** Each sphere represents a microarray sample. The percentage values in the axes parentheses designate proportion of overall variance as described by each PC. PC1 principal component 1 ( $x$ -axis); PC2 principal component 2 ( $y$ -axis); PC3 principal component 3 ( $z$ -axis). PC1 describes the predominant amount of variance (15.6%). Selection

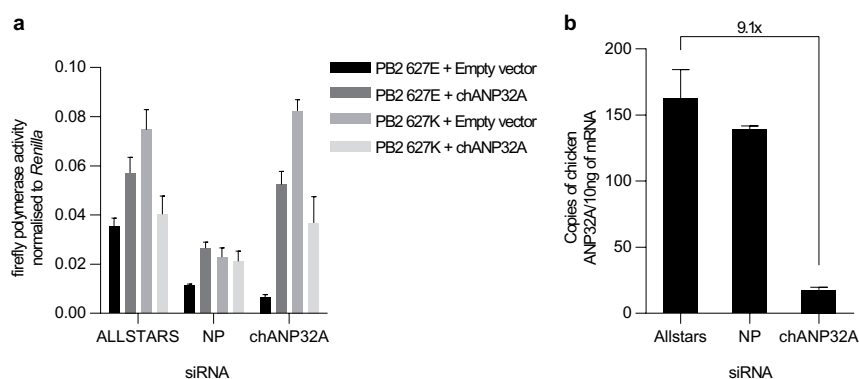
of negative clones (red), parent Wg3H cells (blue) and positive clones: 377 (purple), 386 (orange) and 365 (green) and 476 (cyan) arrays are distinguished by colour, and passage numbers 1 and 12 are distinguished by the size of spheres. Negative arrays are dispersed, while parent cells are accumulated further to the right of PC1 and upwards of PC2. Positive clones show distinct variability in their location while passaging reduced their separation from parent cells. This analysis accompanies Fig. 1.





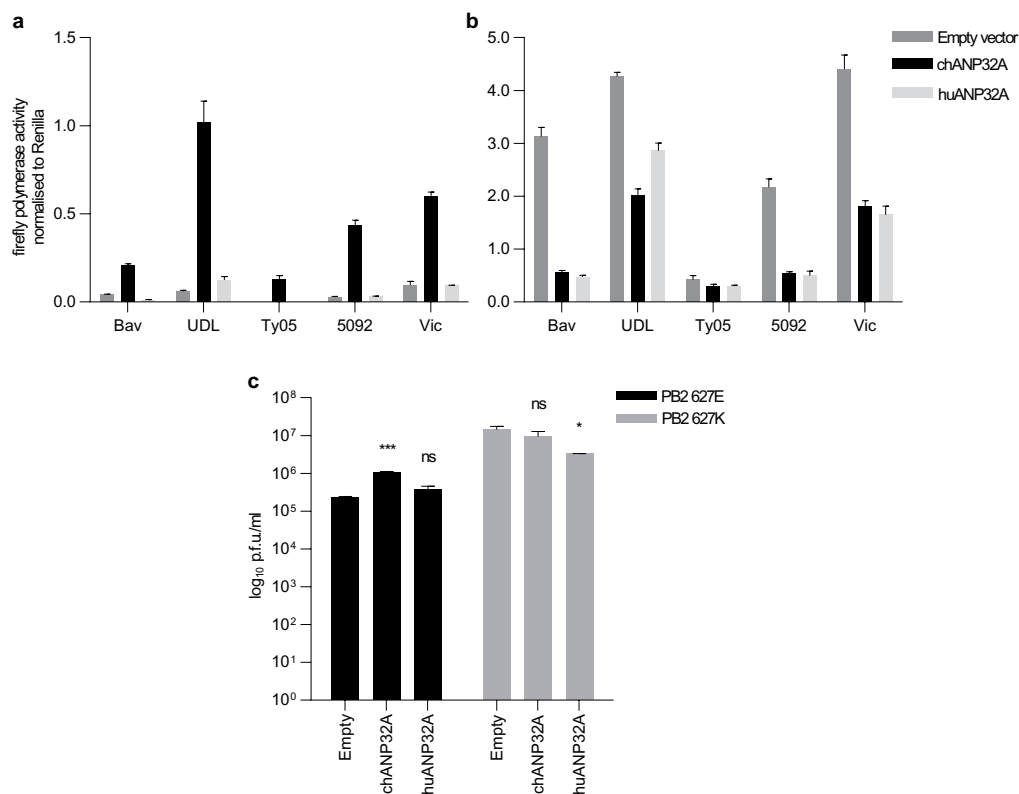
**Extended Data Figure 2 | Confirmation of chANP32A and chANP32B expression in RH clones by qRT-PCR.** RNA was extracted from the RH clones after testing for influenza polymerase activity and analysed by microarray for chicken transcripts. The same RNA was used to validate identification of ANP32A by confirming the level of expression of ANP32A (and ANP32B as control) in the parent Wg3h cells, positive clones, passaged positive clones and a selection of negative clones. **a**, Copy

numbers of chANP32A mRNA were calculated by qRT-PCR against a standard curve generated with chANP32A cDNA using primers specific for chANP32A. **b**, Copy number of ANP32B mRNA were measured by qRT-PCR against a standard curve generated with chANP32B cDNA using primers specific for chANP32B. ( $n = 3$  technical replicates; error as s.e.m.). This analysis accompanies Fig. 1.



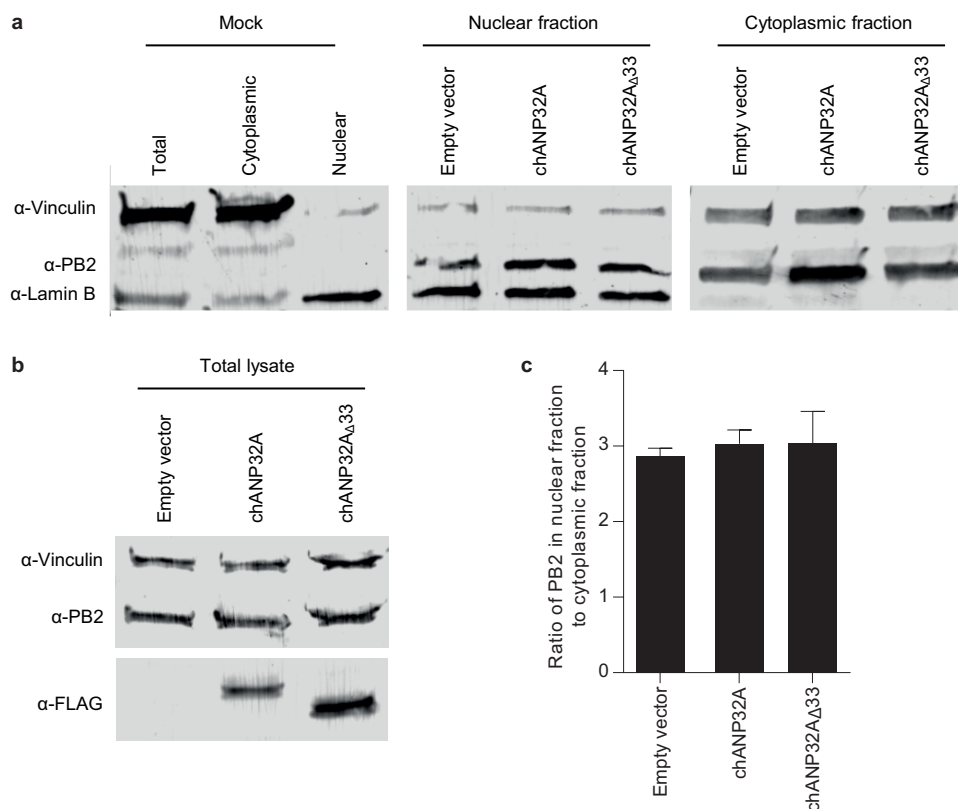
**Extended Data Figure 3 | Knockdown of chANP32A in positive RH clone 476 diminished the ability to support avian influenza polymerase activity.** **a**, Positive RH clone 476 cells were transfected with 100 nM of siRNA targeting NP, chANP32A or no target (Allstars). After 48 h cells were transfected with mouse-polI-firefly minigenome reporter, avian influenza polymerase (H5N1 50–92) with either PB2 627E or 627K, *Renilla* control and either empty plasmid or codon optimised chANP32A (codon

optimization according to algorithm by GeneArt with manual editing). (Data are luciferase activity measured after a further 24 h;  $n = 3$  biological replicates; errors are displayed as s.e.m.). **b**, Knockdown of chANP32A was confirmed by qRT-PCR of RNA extracted from siRNA treated cells, calculated using a standard curve generated with chANP32A cDNA, using primers specific for chANP32A ( $n = 3$  biological replicates; errors are displayed as s.e.m.). This analysis accompanies Fig. 1.



**Extended Data Figure 4 | Expression of chANP32A in human cells permits influenza polymerase activity of several avian influenza polymerases and an avianized human influenza polymerase and increases avian virus replication.** 293T cells were transfected with empty vector, chANP32A or huANP32A. **a**, **b**, 20 h later, cells were transfected with pHOM1-firefly minigenome reporter, and the polymerase set from low pathogenicity avH1N1 (Bav) or H9N2 (UDL), highly pathogenic H5N1 (50-92), H5N1 (Ty05), or huH3N2 (Vic) viruses, with either PB2 627E (**a**) or 627K (**b**) and *Renilla* expression control. After a further 24 h luciferase activity was measured. (Data are mean PB2 627E or K polymerase activity normalized to *Renilla*;  $n = 3$  biological replicates; error

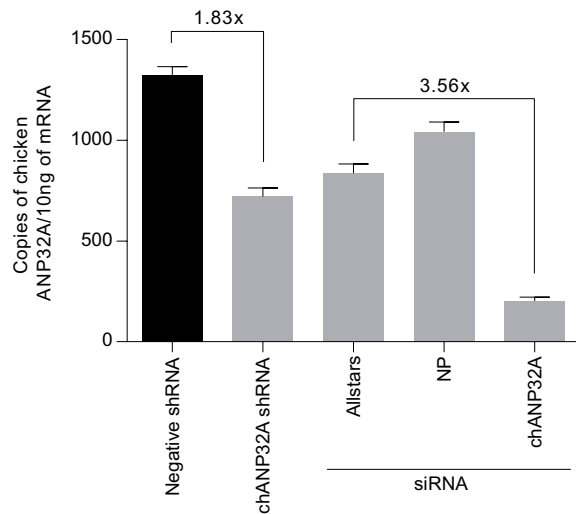
plotted as s.e.m. of the ratio; pattern of results consistent in at least three independent experiments). This analysis accompanies Fig. 1. **c**, 20 h after transfection with ch or huANP32A or empty vector, cells were infected with avian-like influenza virus (H5N1 Ty05:PR8 6:2 recombinant virus) (MOI 0.1) bearing PB2 627E (black bars) or PB2 627K (grey bars). Infected cells were incubated at 37°C and cell supernatant titrated for infectious virus at 24 h post infection on MDCK cells by plaque assay. (Data displayed as log<sub>10</sub> plaque forming units per ml;  $n = 3$  biological replicates; error plotted as s.e.m.; one-way ANOVA, comparisons to empty vector, NS= not significant, \* $P < 0.05$  \*\*\* $P < 0.001$ ; pattern of results consistent in at least three independent experiments). This analysis accompanies Fig. 2.



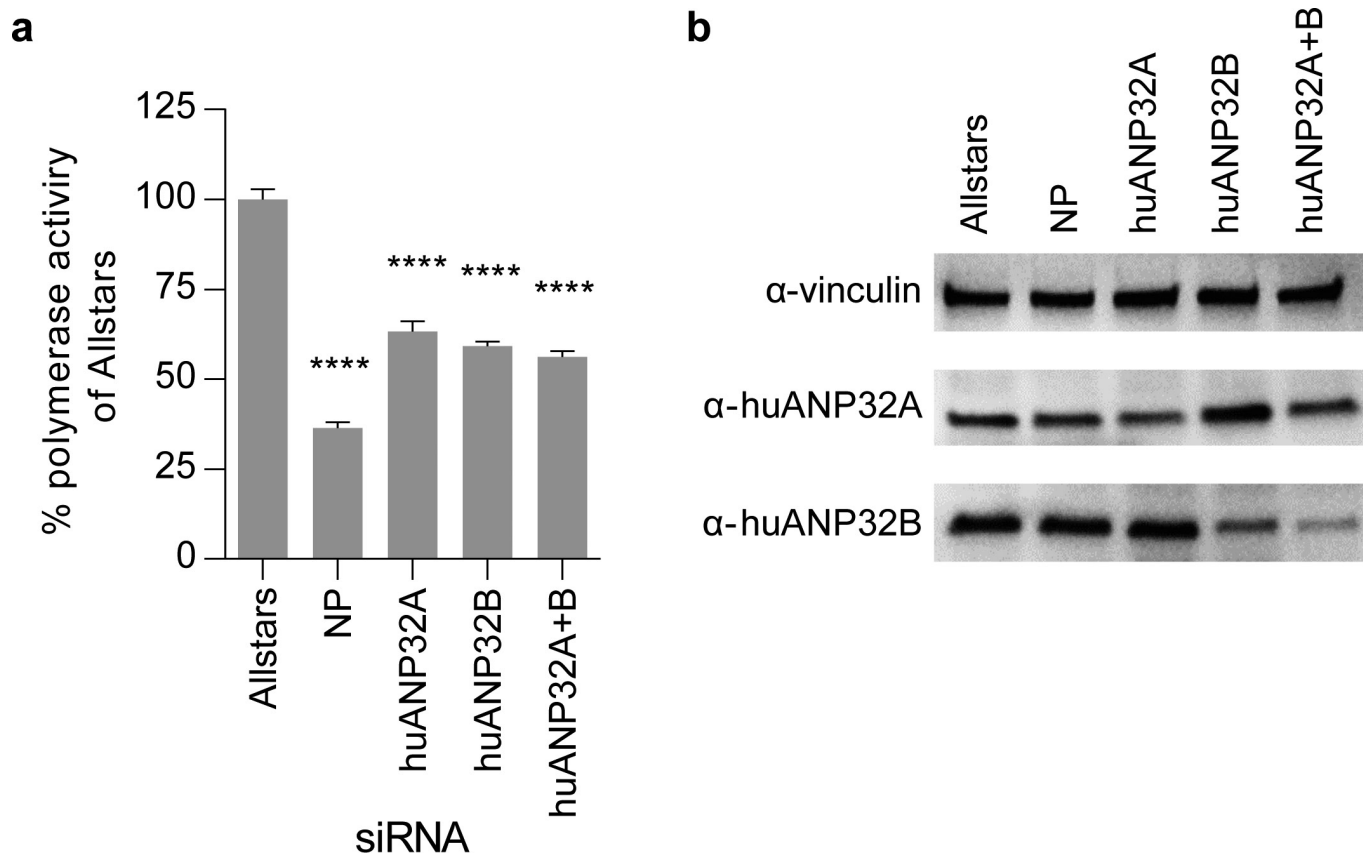
**Extended Data Figure 5 | chANP32A does not alter expression or nuclear accumulation of avian PB2 protein in human cells.** 293T cells were transfected with pHOM1-firefly minigenome, avian influenza polymerase and NP of H5N1 50–92 (PB2 627E) together with empty vector, chANP32A or chANP32A $\Delta$ 33 or cells were untransfected (Mock). Cell monolayers were harvested after 24 h and lysed in 0.1% NP40 lysate buffer and total fractions taken before centrifugation to generate a nuclear pellet and cytoplasmic fraction. Nuclear pellets were resuspended in 1% NP40 buffer. **a**, Protein levels of vinculin (cytoplasmic marker) and lamin B (nuclear marker) and of PB2 in total, nuclear or cytoplasmic fractions

were analysed by immunoblotting. **b**, Total lysates were immunoblotted for vinculin, PB2 and Flag peptide. **c**, Immunoblots were quantified using Image Studio Lite V5.2. The ratio of nuclear to cytoplasmic PB2 was calculated by dividing the ratio of PB2 to vinculin by the ratio of PB2 to lamin B from the cytoplasmic and nuclear fractions, respectively. Data are the mean ratios from three independent experiments (excepting chANP32A $\Delta$ 33 for which only 2 data points were available), error bars are s.e.m. Data are not statistically significantly different by one-way ANOVA. This analysis accompanies Fig. 2.



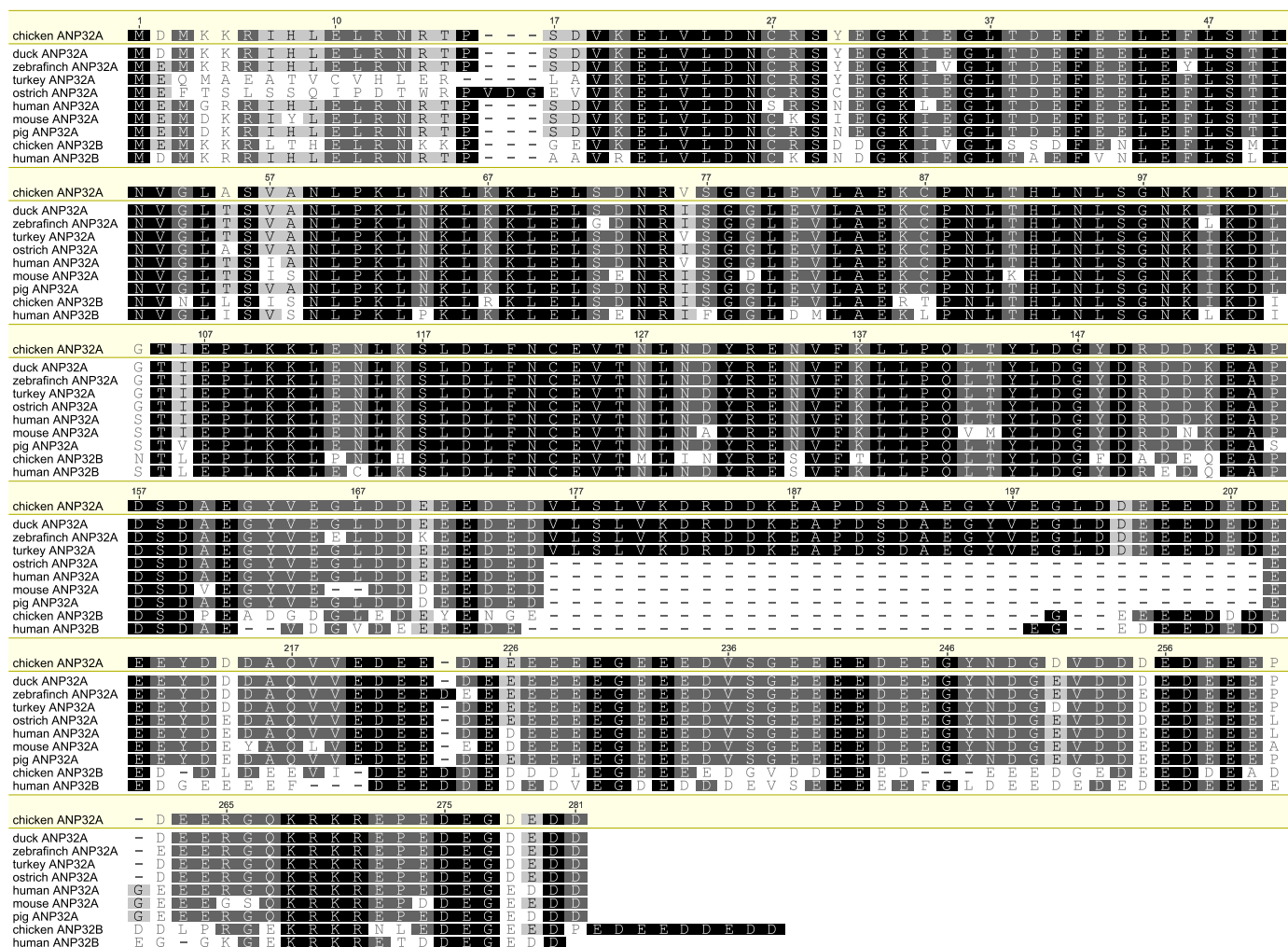


**Extended Data Figure 6 | Quantification of knockdown of chANP32A in chicken cells.** DF-1 cells were transduced with VSV-G lentiviral vectors that delivered a transgene expressing shRNA directed against chANP32A or a negative sequence and the puromycin gene. Puromycin selected cells were transfected with siRNA (100 nM) (underlined). RNA was extracted from untreated shRNA cells and siRNA-treated shRNA cells. Knockdown of chANP32A was quantified by qRT-PCR of the extracted RNA, calculated using a standard curve generated with chANP32A cDNA, using primers specific for chANP32A. Fold decrease of RNA copies is displayed compared to negative shRNA DF-1 or ALLstars treated chANP32A shRNA DF-1 cells. ( $n = 3$  biological replicates; error displayed as s.e.m.). This analysis accompanies experiments in Fig. 3a–c.



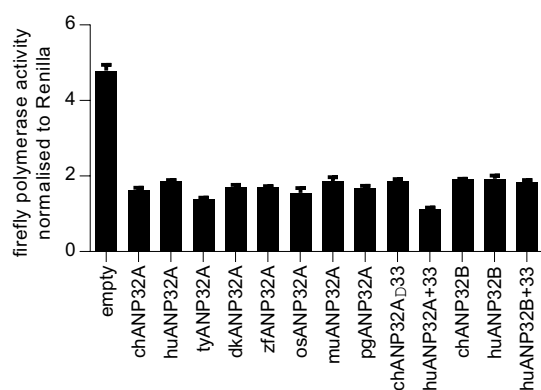
**Extended Data Figure 7 | siRNA knockdown demonstrates that human-adapted influenza polymerase activity is dependent on huANP32A and huANP32B in human cells.** **a**, 293T cells were transfected with siRNA (100 nM) against NP, huANP32A, huANP32B or both huANP32A and huANP32B (50 nM each). After 48 h, cells were transfected with pHOM1-firefly minigenome, human-adapted avian influenza polymerase (H5N1 50–92 PB2 627K), and *Renilla* expression control. Luciferase activity

was measured after a further 24 h. (Data are firefly activity normalized to *Renilla*, plotted as % of Allstars;  $n = 3$  biological replicates; error as s.e.m.; one-way ANOVA comparisons to Allstars, \*\*\*\* $P < 0.0001$ ); **b**, Knockdown of gene targets was verified by immunoblotting using antibody against vinculin, huANP32A and huANP32B. This analysis accompanies Fig. 3e, f.



**Extended Data Figure 8 | Alignment of ANP32A proteins reveals significant homology except for an extra 33 amino acid sequence in birds that is absent in mammals and ostrich and lacking from ANP32B family members.** The protein sequences of ANP32A for chicken, duck, zebra finch, turkey, ostrich, human, mouse and pig together with sequences of ANP32B for chicken and human were aligned using

Geneious R6 software. chANP32A is set as the reference sequence, and colours represent similarity of amino acid identity (black, 100%; dark grey, 80–100%; light grey, 60–80%; white, <60%). Gaps are annotated by dashes. Residue numbers correspond to chANP32A. The 33 amino acid sequence found in avian species is situated between residues 176–208. This analysis accompanies Fig. 4.



**Extended Data Figure 9 | Expression of ANP32A and B proteins reduced human-adapted influenza polymerase activity in human cells.** 293T cells were transfected with Flag-tagged ANP32 constructs and after 20 h transfected with pHOM1-firefly minigenome reporter, human-adapted influenza polymerase (H5N1 50–92 with PB2 627K, together with *Renilla* expression control. Cells were assayed for luciferase activity 24 h later. (Data are PB2 627K polymerase activity normalized to *Renilla*;  $n = 3$  biological replicates; error plotted as s.e.m. of the ratio; one-way ANOVA, all constructs were significantly reduced compared to empty vector ( $P < 0.0001$ ); pattern of results consistent in at least three independent experiments). These data relate to Fig. 4.



**Extended Data Table 1 | Genes common between the avian influenza polymerase positive radiation hybrid cells**

	Gene Name	Chromosome
Venn 1	IQSEC3	1
	CCDC77	1
	WNK1	1
	KDM5A	1
	MAP2K1	10
	ZWILCH	10
	PIAS1	10
	FEM1B	10
	GLCE	10
	KIF23	10
	SPG11	10
	TIPIN	10
	RPL4	10
	CLN6	10
	ANP32A	10
	MORF4L1	10
	EIF3J	10
	CASC4	10
	BLM	10
	NUB1	2
	Unannotated	2
	RBM33	2
	WDR48	2
	PAXIP1	2
	Unannotated	4
	HPRT1	4
	DDX26B	4
	SLC9A6	4
	HTATSF1	4
	LOC422249	4
	FAM122B	4
	MOSPD1	4
	MMGT1	4
	MIR1726	6
	LOC430470	Uncharacterised
Venn 2	MAP2K1	10
	PIAS1	10
	FEM1B	10
	GLCE	10
	KIF23	10
	SPG11	10
	TIPIN	10
	CLN6	10
	ANP32A	10
	EIF3J	10
	CASC4	10
	BLM	10

Microarray analysis of RNA extracted from RH clones revealed the expression of chicken genes. The gene list from Venn 1 shows common genes between the P1 positive hybrid clones when compared to parent Wg3h hamster cells. The gene list from Venn 2 shows common genes between P12 positive hybrid clones when compared to parent and 365P1 (positive RH clone) vs 365P12 (reverted RH clone). This analysis accompanies the microarray analysis Venn diagrams in Fig. 1b and c.

# A *LAIR1* insertion generates broadly reactive antibodies against malaria variant antigens

Joshua Tan<sup>1,2,3\*</sup>, Kathrin Pieper<sup>1\*</sup>, Luca Piccoli<sup>1\*</sup>, Abdirahman Abdi<sup>2</sup>, Mathilde Foglierini<sup>1</sup>, Roger Geiger<sup>1,4</sup>, Claire Maria Tully<sup>2</sup>, David Jarrossay<sup>1</sup>, Francis Maina Ndungu<sup>2</sup>, Juliana Wambua<sup>2</sup>, Philip Bejon<sup>2,3</sup>, Chiara Silacci Fregni<sup>1</sup>, Blanca Fernandez-Rodriguez<sup>1</sup>, Sonia Barbieri<sup>1</sup>, Siro Bianchi<sup>5</sup>, Kevin Marsh<sup>2,3</sup>, Vandana Thathy<sup>2</sup>, Davide Corti<sup>5</sup>, Federica Sallusto<sup>1</sup>, Peter Bull<sup>2,3§</sup> & Antonio Lanzavecchia<sup>1,4§</sup>

*Plasmodium falciparum* antigens expressed on the surface of infected erythrocytes are important targets of naturally acquired immunity against malaria, but their high number and variability provide the pathogen with a powerful means of escape from host antibodies<sup>1–4</sup>. Although broadly reactive antibodies against these antigens could be useful as therapeutics and in vaccine design, their identification has proven elusive. Here we report the isolation of human monoclonal antibodies that recognize erythrocytes infected by different *P. falciparum* isolates and opsonize these cells by binding to members of the RIFIN family. These antibodies acquired broad reactivity through a novel mechanism of insertion of a large DNA fragment between the V and DJ segments. The insert, which is both necessary and sufficient for binding to RIFINs, encodes the entire 98 amino acid collagen-binding domain of LAIR1, an immunoglobulin superfamily inhibitory receptor encoded on chromosome 19. In each of the two donors studied, the antibodies are produced by a single expanded B-cell clone and carry distinct somatic mutations in the LAIR1 domain that abolish binding to collagen and increase binding to infected erythrocytes. These findings illustrate, with a biologically relevant example, a novel mechanism of antibody diversification by interchromosomal DNA transposition and demonstrate the existence of conserved epitopes that may be suitable candidates for the development of a malaria vaccine.

To identify individuals who may produce antibodies that broadly react with *P. falciparum*-infected erythrocytes (IEs), we developed an improved mixed agglutination assay (Fig. 1a). Plasma from adults ( $n = 557$ ) living in a malaria-endemic region in Kilifi, Kenya, were initially tested in pools of five (Fig. 1b) and then individually for their capacity to agglutinate mixtures of erythrocytes infected with three culture-adapted Kenyan parasite isolates, each stained with a different DNA dye. Most plasma samples formed single-colour agglutinates, but three were able to form mixed-colour agglutinates with at least six isolates (Fig. 1c).

From two selected donors (C and D) whose plasma formed mixed agglutinates with eight parasite isolates, we immortalized immunoglobulin G (IgG)<sup>+</sup> memory B cells<sup>5</sup> and screened the culture supernatants for the capacity to stain erythrocytes infected with the eight isolates. Surprisingly, most antibodies isolated from these donors stained multiple isolates, with the best antibodies, such as MGC34, MGD21 and MGD39, recognizing all eight isolates tested (Fig. 1d). Conversely, a few antibodies, such as MGD13, were specific for a single isolate. In all cases, only a fraction of IEs were stained (Fig. 1e) and this fraction varied with different antibodies, possibly reflecting different clonal expression of the relevant antigen. Overall, these findings show that broadly reactive antibodies against IEs can be generated in response to malaria infection.

We investigated the molecular basis of the broad antibody reactivity by comparing the sequences of the antibodies isolated from the two donors. While the antibodies with narrow reactivity showed classical VDJ organization of the heavy (H) chain gene, all the broadly reactive antibodies (14 from donor C, 13 from donor D) carried a large insert of more than 100 amino acids between their V and DJ segments (Fig. 2a and Extended Data Figs 1–3). In both donors, the core of the inserts encoded an amino acid sequence that was 85–96% identical to the extracellular domain of LAIR1, a collagen-binding inhibitory receptor encoded in the leukocyte receptor locus on chromosome 19 (ref. 6). However, in each donor, the broadly reactive antibodies used a distinct VH/JH combination (VH3-7/JH6 in donor C and VH4-4/JH6 in donor D) and had junctions of distinct length between the V, LAIR1 and J segments. In addition, the broadly reactive antibodies from donor D shared a single light (L) chain (VK1-8/JK5), while the antibodies from donor C had one of three different L chains (VK1-5/JK2, VK4-1/JK2, VL7-43/JL3) (Extended Data Fig. 4). All the broadly reactive antibodies carried a high load of somatic mutations spanning the whole V–LAIR1–DJ region. The mutations in the VH segment were used to reconstruct genealogy trees showing a developmental pathway with progressive acquisition of somatic mutations (Fig. 2b, c). Notably, the trees were consistent with those generated using only the LAIR1 insert or the VL sequence (Extended Data Fig. 5). These findings indicate that, within each individual, a single B-cell clone carrying a LAIR1 insert expanded after stimulation by malaria antigens and progressively accrued mutations in the LAIR1, VH and VL regions.

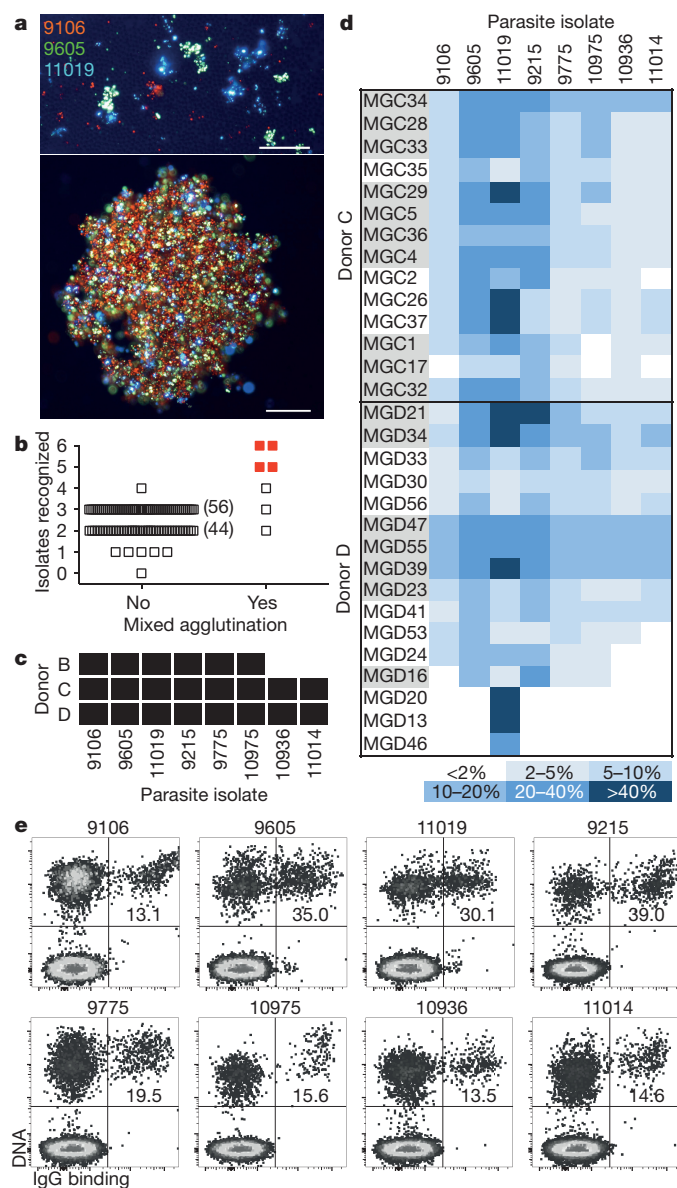
To explore the mechanism that led to the generation of the LAIR1-containing antibodies, we compared complementary DNA and genomic DNA sequences obtained from the antibody-producing B-cell clones (Fig. 2d). In both donors, the genomic DNA contained a *LAIR1* insert that was larger than that found in the corresponding cDNA. In particular, in donor C, the insert comprised not only the 294 base pair (bp) exon encoding the extracellular LAIR1 domain, but also a 190 bp 5' intronic region of the *LAIR1* gene that was partially spliced out in the messenger RNA, and a shorter 23 bp 3' intronic region that was maintained in the mRNA (Extended Data Fig. 6a). Donor D had a somewhat different genomic insertion, with larger 5' (378 bp) and 3' (60 bp) *LAIR1* intronic sequences, and, 5' of the *LAIR1* insertion, an additional sequence of 135 bp corresponding to an intergenic sequence of chromosome 13 (Extended Data Fig. 6b, c). In this donor, the entire *LAIR1* 5' intronic sequence and much of the 5' chromosome 13 sequence were spliced out in the mRNA (Extended Data Fig. 6d). The spliced intronic *LAIR1* region contained a duplicated 135 bp element with a very high load of somatic mutations (Extended Data Fig. 6e).

The finding that the inserts were located exactly between the V and DJ segments and were joined to these segments by N nucleotides

<sup>1</sup>Institute for Research in Biomedicine, Università della Svizzera Italiana, Via Vincenzo Vela 6, 6500 Bellinzona, Switzerland. <sup>2</sup>KEMRI-Wellcome Trust Research Programme, CGMRC, PO Box 230, 80108 Kilifi, Kenya. <sup>3</sup>Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK. <sup>4</sup>Institute for Microbiology, ETH Zurich, Wolfgang-Pauli-Strasse 10, 8093 Zurich, Switzerland. <sup>5</sup>Humabs BioMed SA, 6500 Bellinzona, Switzerland.

\*These authors contributed equally to this work.

§These authors jointly supervised this work.



**Figure 1 | Identification of broadly reactive monoclonal antibodies against IEs.** **a**, Fluorescence microscopy images of single agglutinates (top) and a triple agglutinate (bottom). Scale bar, 50  $\mu$ m. **b**, **c**, Plasma (pooled in groups of five) from immune adults were screened against six parasite isolates using the triple mixed agglutination assay (**b**). Pools that formed mixed agglutinates with at least five isolates (in red) were further investigated for individual reactivity against an extended panel of eight isolates (**c**). **d**, Heat map showing the percentage of IEs of eight parasite isolates stained by monoclonal antibodies isolated from two donors ( $n = 1$ ). Closely related antibodies are grouped in alternating colours. **e**, Example of staining of IEs by the broadly reactive antibody MGD25.

would suggest that RAG might be involved in the insertion process. Indeed, cryptic recombination signal sequences (RSSs) that followed the 12/23 rule were found flanking both the *LAIR1* and the chromosome 13 inserts, although their RSS prediction scores were low and they were not positioned precisely at the ends of the inserts (Extended Data Fig. 7). As RAG acts by excising a target DNA sequence, we investigated whether, in B cells making *LAIR1*-containing antibodies, one of the two *LAIR1* alleles on chromosome 19 would be deleted. By sequencing genomic DNA from T cells of donor C, we identified a heterozygous nucleotide site in the chromosome 19 *LAIR1* exon sequence. Surprisingly, both alleles were also present in the B cells producing *LAIR1*-containing antibodies

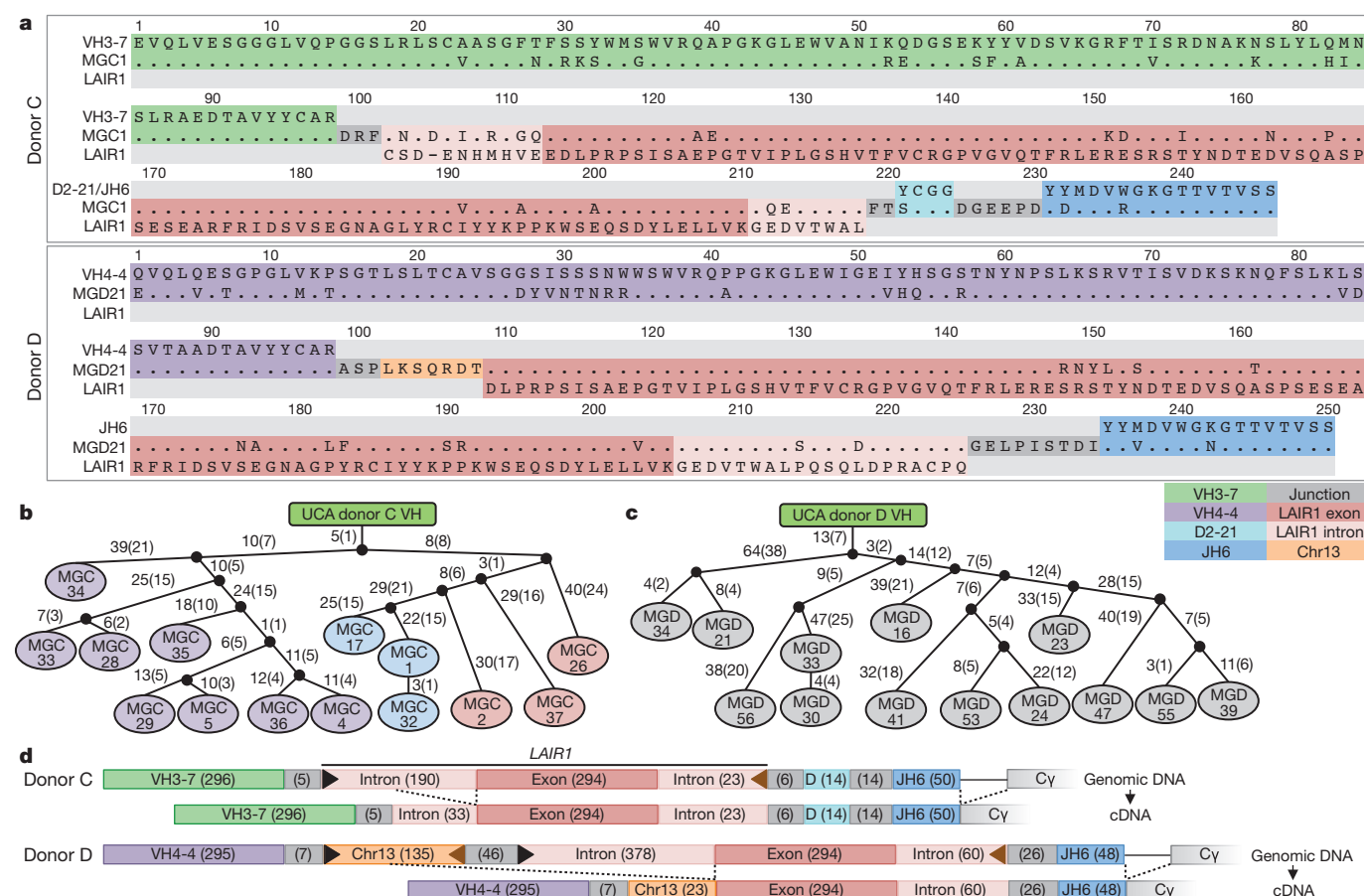
(Extended Data Fig. 8), a finding that is inconsistent with the 'cut-and-paste' function of RAG.

To determine the contribution of the mutated VH, VL and *LAIR1* domains to the antibody specificity, we generated a panel of constructs and fusion proteins based on the broadly reactive antibody MGD21 (Fig. 3a). Substitution of the V, J or L chain of MGD21 with that of an unrelated antibody did not affect binding to IEs (Fig. 3b), suggesting that these elements are dispensable for binding. In contrast, deletion of the *LAIR1* insert, or its reversion to the unmutated genomic sequence, led to a complete loss of binding. Furthermore, fusion proteins displaying only the mutated *LAIR1* domain bound to IEs, although with lower affinity. To dissect the contribution of the somatic mutations of the *LAIR1* insert to antigen binding, we created a set of *LAIR1*-Fc fusion proteins carrying, in various combinations, the mutations shared by MGD21 with other antibodies of the same clonal family. We tested the mutants for binding to IEs and to collagen, which is the natural ligand of *LAIR1*. Interestingly, two distinct kinds of mutations were identified: those that reduced collagen binding (P106S and P107R) and those that increased binding to IEs (T67L, N69S and A77T) (Fig. 3c). Collectively, these findings indicate that the binding of the broadly reactive antibodies to IEs relies mainly on the mutated *LAIR1* domain, which evolves under selective pressure to lose collagen binding and gain binding to IEs.

To identify the antigen(s) recognized by the *LAIR1*-containing antibodies, we generated stable *P. falciparum* 3D7 lines that were enriched (3D7-MGD21<sup>+</sup>) or depleted (3D7-MGD21<sup>-</sup>) of MGD21 reactivity (Extended Data Fig. 9a). Western blot analysis showed two specific MGD21-reactive bands of 40–45 kilodaltons (kDa) in erythrocyte ghosts and in MGD21 immunoprecipitates prepared from 3D7-MGD21<sup>+</sup> IEs (Fig. 4a). Analysis of the MGD21 immunoprecipitates by liquid chromatography coupled with mass spectrometry (LC-MS) revealed that a member of the A-type RIFIN family (PF3D7\_1400600) was significantly enriched in 3D7-MGD21<sup>+</sup> immunoprecipitates as compared to 3D7-MGD21<sup>-</sup> immunoprecipitates ( $\log_2$  fold change >2;  $P < 0.01$ ) (Fig. 4b). PF3D7\_1400600 and a second A-type RIFIN (PF3D7\_1040300) were also identified in 3D7-MGD21<sup>+</sup> but not in 3D7-MGD21<sup>-</sup> ghosts in the absence of immunoprecipitation (Extended Data Fig. 9b). In contrast, four other RIFINs, including one recently characterized for its capacity to induce rosetting (PF3D7\_0100400)<sup>3</sup>, were detected in similar amounts in both 3D7-MGD21<sup>+</sup> and 3D7-MGD21<sup>-</sup> ghosts. We found that enrichment for 3D7-MGD21<sup>+</sup> IEs greatly increased recognition by all the other broadly reactive antibodies from donor D tested and, notably, by two broadly reactive antibodies from donor C (Extended Data Fig. 9c), suggesting that these antibodies recognize the same antigens. Similar results were obtained with the Kenyan isolate 9605 (Extended Data Fig. 9d, e).

The binding of the *LAIR1*-containing antibodies to specific RIFINs was confirmed by the finding that MGD21 stained CHO cells transfected with the candidate antigens PF3D7\_1400600 and PF3D7\_1040300, but not with irrelevant RIFINs that were similarly expressed (PF3D7\_0100400 and PF3D7\_0100200) or not detected (PF3D7\_1100500) in 3D7-MGD21<sup>+</sup> and 3D7-MGD21<sup>-</sup> ghosts (Fig. 4c). Furthermore, MGD21 and an Fc fusion protein containing the MGD21 *LAIR1* domain stained CHO cells transfected with a RIFIN chimera containing the constant region of PF3D7\_0100200 and the variable region of PF3D7\_1400600, but not cells transfected with the inverse chimera (Extended Data Fig. 9f, g), indicating that MGD21 binds to the variable region. Collectively, these results indicate that the *LAIR1*-containing antibodies recognize specific members of the RIFIN family in different *P. falciparum* isolates.

Addition of MGD21 to 3D7 culture did not interfere with parasite growth and did not result in decreased expression of the antigen(s) (Extended Data Fig. 9h, i). In addition, when tested in a rosette inhibition assay with O<sup>+</sup> or A<sup>+</sup> erythrocytes, MGD21 did not show a consistent inhibitory effect ( $P > 0.1$  for both blood groups)



**Figure 2 | Broadly reactive antibodies contain a mutated LAIR1 insert and are produced by expanded clones.** **a**, Protein sequence alignment of MGC1 and MGD21 with germline-encoded sequences of the corresponding VH (green or purple), DH (cyan), JH (blue) and LAIR1 segments (exon in red and intronic sequences in light red). Chromosome 13 (Chr13) sequences are shown in orange while grey areas show junctional sequences for which no homology was found. **b**, **c**, Genealogy trees drawn from the VH nucleotide sequences of antibodies from donors

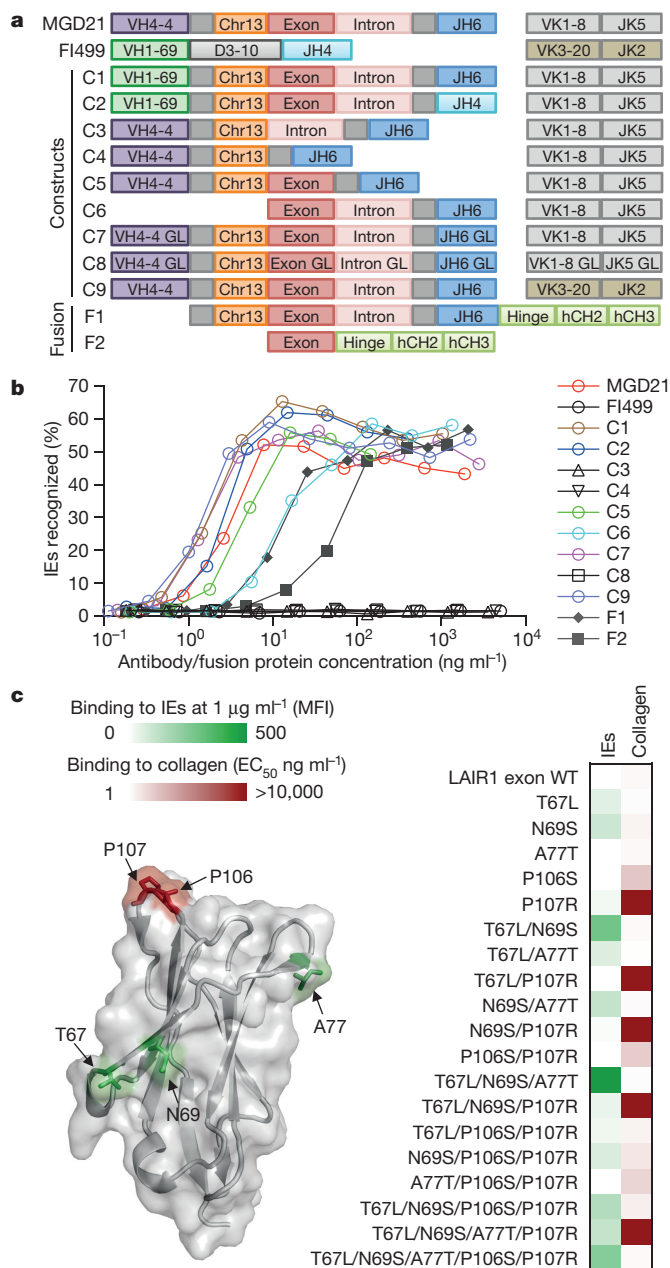
C (**b**) and D (**c**). In the donor C genealogy tree, antibodies that use different light chains are highlighted in different colours. Shown are the nucleotide and amino acid substitutions, with the latter in parentheses. UCA, unmutated common ancestor. **d**, Scheme showing genomic DNA and cDNA of LAIR1-containing antibodies from donors C and D. Shown are the lengths of the fragments (bp in parentheses), cryptic 12 and 23 RSS sites (black and brown triangles, respectively) and splicing positions (dashed lines).

(Extended Data Fig. 9j). In contrast, MGD21, as well as MGC34, could agglutinate erythrocytes infected with 3D7 or the Kenyan isolate 11019 (Extended Data Fig. 9k). Furthermore, MGD21 showed a strong capacity to opsonize 3D7 IEs for phagocytosis by human monocytes (Fig. 4d). Opsonization was dependent on an intact Fc, as a mutant lacking Fc receptor binding (MGD21 LALA) did not induce phagocytosis. Similar results were obtained with other broadly reactive antibodies isolated from both donors and with a different parasite isolate (11019) (Extended Data Fig. 9l), suggesting that these broadly reactive antibodies could be effective in promoting phagocytosis and destruction of IEs *in vivo*.

Our study opens several questions as to the potential use of RIFINs as targets for passive and active vaccination. RIFINs represent the largest family (~150 genes) of variant antigens expressed on IEs, some of which have been implicated in severe malaria<sup>3</sup>. The LAIR1-containing antibodies have potent agglutinating and opsonizing activity, which would be consistent with their role in decreasing the burden of IEs *in vivo* by enhancing parasite clearance. However, the staining of only a fraction of IEs by the LAIR1-containing antibodies is consistent with the clonal expression of RIFINs<sup>3</sup> and suggests that these antibodies may not be sufficient to take full control of the infection. It will be interesting to determine whether the LAIR1-containing antibodies recognize RIFINs that are expressed at other stages of the parasite life cycle, such as sporozoites, merozoites and gametocytes<sup>7,8</sup>, which may create new opportunities for vaccine design.

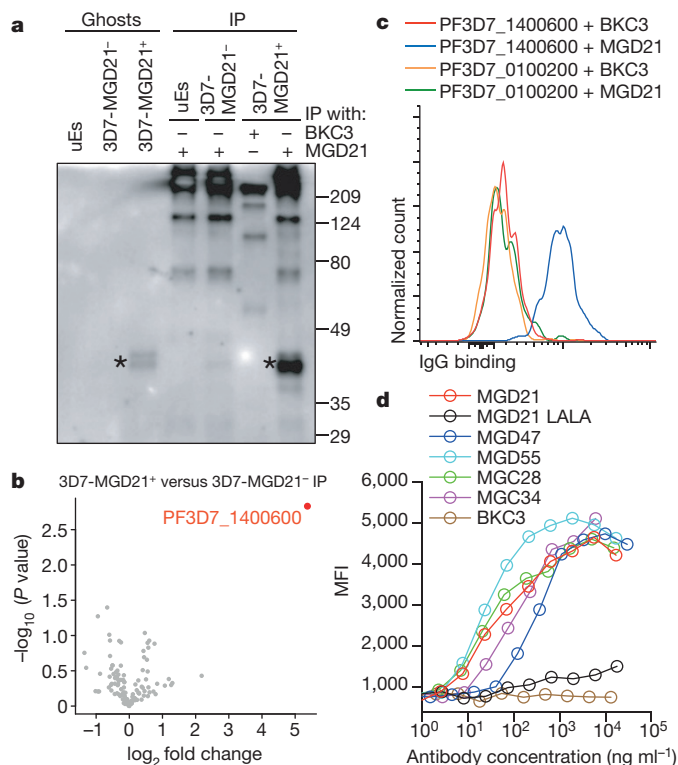
The unusual architecture of the LAIR1-containing antibodies illustrates a novel mechanism of interchromosomal DNA transposition that can contribute to antibody diversification (Extended Data Fig. 10). The precise location of the LAIR1 and chromosome 13 inserts between the V and DJ segments, as well as the presence of N nucleotides and cryptic 12/23 RSSs at the ends of the inserts, would be compatible with a role for the RAG enzyme. RAG has been implicated in interchromosomal genomic rearrangements at cryptic RSSs outside the immunoglobulin and T-cell antigen receptor (TCR) loci<sup>9,10</sup>, and in the formation of chromosomal translocations found in human lymphomas<sup>11</sup>. However, RSSs are frequently found in the genome and are generally inactive, according to recent data<sup>12,13</sup>. Furthermore, the conservation of the two LAIR1 alleles in B cells producing LAIR1-containing antibodies is inconsistent with a RAG-mediated 'cut-and-paste' pathway and suggests a new mechanism by which LAIR1 DNA is duplicated. This mechanism may involve reverse transcription of pre-mRNA and subsequent insertion of the duplicated fragment to repair a DNA double-strand break, as recently proposed<sup>14</sup>. It is also possible that gene conversion<sup>15</sup> or AID-dependent genomic instability caused by chronic *Plasmodium* infection<sup>16</sup> may contribute to the production of LAIR1-containing antibodies. AID can lead to insertions and deletions of multiple codons in the V genes, which contribute to the specificity of the antibody in the context of the whole V gene<sup>17,18</sup>. Nevertheless, to the best of our knowledge, these insertions are distributed over the whole V-gene sequence, are of smaller size and cannot be traced back to a particular genomic sequence as in the case of LAIR1.





**Figure 3 | The mutated LAIR1 insert is necessary and sufficient for binding to IEs.** **a**, Design of modified MGD21 antibody constructs with selected regions replaced with counterparts from an unrelated antibody (FI499) (C1–C2, C9), deleted (C3–C6), or reverted to germline (GL) (C7–C8). Fc fusion proteins that incorporated the LAIR1 insert, junction and downstream sequences (F1), as well as the LAIR1 domain alone (F2), were also designed. **b**, Binding of MGD21 constructs and Fc fusion proteins to IEs (representative of  $n = 2$  independent experiments). **c**, Selected amino acid substitutions found in MGD21 were added individually or in different combinations to the germline LAIR1–Fc fusion protein. These mutants were tested for binding to collagen and to IEs. Shown are the effects of the mutations on binding to IEs or collagen (one representative of  $n = 2$  independent experiments) and their location on the LAIR1 structure<sup>19</sup> (Protein Data Bank accession 3KGR). Gain of IE binding is shown in green (background mean fluorescence intensity (MFI) values subtracted). Loss of collagen binding (half-maximum effective concentration (EC<sub>50</sub>) enzyme-linked immunosorbent assay (ELISA) values) is shown in red. WT, wild type.

The transposition of *LAIR1* (and chromosome 13) sequences into V–DJ genes is the first example of an insertion that gives rise to a functional antibody in which the insert represents the fundamental binding



**Figure 4 | LAIR1-containing antibodies bind to distinct RIFINs and opsonize IEs.** **a**, Western blot showing MGD21 binding to erythrocyte ghosts and MGD21 immunoprecipitates (IP) prepared from 3D7-MGD21<sup>+</sup> and 3D7-MGD21<sup>-</sup> IEs (representative of  $n = 2$  independent experiments). Controls include uninfected erythrocytes (uEs) and immunoprecipitates with an irrelevant antibody (BKC3). Specific bands are marked with asterisks. Anti-human IgG was used as the secondary antibody, resulting in detection of antibodies used for immunoprecipitation alongside antigens of interest. For gel source data, see Supplementary Fig. 1. Numbers on right indicate kDa. **b**, Volcano plot from LC-MS analysis of MGD21 immunoprecipitates prepared from 3D7-MGD21<sup>+</sup> IEs versus from 3D7-MGD21<sup>-</sup> IEs (from  $n = 4$  independent experiments). Statistical significance was evaluated by Welch tests ( $P < 0.01$  for PF3D7\_1400600). **c**, MGD21 and BKC3 staining of CHO cells transfected with a specific (PF3D7\_1400600) or an irrelevant (PF3D7\_0100200) RIFIN (representative of  $n = 5$  independent experiments). **d**, Opsonic phagocytosis of 3D7-MGD21<sup>+</sup> IEs by monocytes ( $n = 3$  for MGD21, MGD21 LALA and BKC3,  $n = 2$  for others). The IEs were stained with 4',6-diamidino-2-phenylindole (DAPI), which was quantified in monocytes as a measure of phagocytosis. MGD21 LALA is a mutant of MGD21 lacking Fc receptor binding.

element. It remains to be established how often this novel mechanism may give rise to functional antibodies and whether sequences other than *LAIR1* are transposed into immunoglobulin genes. We anticipate that *LAIR1*-containing antibodies will be frequently found in malaria-endemic regions and speculate that the transposed *LAIR1* domain may serve to bind other foreign antigens and possibly also collagen in patients with rheumatic diseases.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 28 August; accepted 16 November 2015.**

**Published online 23 December 2015; corrected online 6 January 2016 (see full-text HTML version for details).**

- Chan, J.-A., Fowkes, F. J. I. & Beeson, J. G. Surface antigens of *Plasmodium falciparum*-infected erythrocytes as immune targets and malaria vaccine candidates. *Cell. Mol. Life Sci.* **71**, 3633–3657 (2014).
- Scherf, A., Lopez-Rubio, J. J. & Riviere, L. Antigenic variation in *Plasmodium falciparum*. *Annu. Rev. Microbiol.* **62**, 445–470 (2008).

3. Goel, S. *et al.* RIFINs are adhesins implicated in severe *Plasmodium falciparum* malaria. *Nature Med.* **21**, 314–317 (2015).
4. Bull, P. C. *et al.* Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. *Nature Med.* **4**, 358–360 (1998).
5. Traggiai, E. *et al.* An efficient method to make human monoclonal antibodies from memory B cells: potent neutralization of SARS coronavirus. *Nature Med.* **10**, 871–875 (2004).
6. Meyaard, L. The inhibitory collagen receptor LAIR-1 (CD305). *J. Leukoc. Biol.* **83**, 799–803 (2008).
7. Le Roch, K. G. *et al.* Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* **301**, 1503–1508 (2003).
8. Florens, L. *et al.* A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419**, 520–526 (2002).
9. Messier, T. L., O'Neill, J. P., Hou, S.-M., Nicklas, J. A. & Finette, B. A. *In vivo* transposition mediated by V(D)J recombinase in human T lymphocytes. *EMBO J.* **22**, 1381–1388 (2003).
10. Vaandrager, J. W., Schuurin, E., Philippo, K. & Kluin, P. M. V. V(D)J recombinase-mediated transposition of the *BCL2* gene to the *IGH* locus in follicular lymphoma. *Blood* **96**, 1947–1952 (2000).
11. Küppers, R., Klein, U., Hansmann, M. L. & Rajewsky, K. Cellular origin of human B-cell lymphomas. *N. Engl. J. Med.* **341**, 1520–1529 (1999).
12. Teng, G. *et al.* RAG represents a widespread threat to the lymphocyte genome. *Cell* **162**, 751–765 (2015).
13. Hu, J. *et al.* Chromosomal loop domains direct the recombination of antigen receptor genes. *Cell* **163**, 947–959 (2015).
14. Onozawa, M. *et al.* Repair of DNA double-strand breaks by templated nucleotide sequence insertions derived from distant regions of the genome. *Proc. Natl Acad. Sci. USA* **111**, 7729–7734 (2014).
15. Reynaud, C.-A., Aoufouchi, S., Faili, A. & Weill, J.-C. What role for AID: mutator, or assembler of the immunoglobulin mutasome? *Nature Immunol.* **4**, 631–638 (2003).
16. Robbiani, D. F. *et al.* *Plasmodium* infection promotes genomic instability and AID-dependent B cell lymphoma. *Cell* **162**, 727–737 (2015).
17. Wilson, P. C. *et al.* Somatic hypermutation introduces insertions and deletions into immunoglobulin V genes. *J. Exp. Med.* **187**, 59–70 (1998).
18. Kepler, T. B. *et al.* Immunoglobulin gene insertions and deletions in the affinity maturation of HIV-1 broadly reactive neutralizing antibodies. *Cell Host Microbe* **16**, 304–313 (2014).
19. Brondijk, T. H. C. *et al.* Crystal structure and collagen-binding site of immune inhibitory receptor LAIR-1: unexpected implications for collagen binding by platelet receptor GPVI. *Blood* **115**, 1364–1373 (2010).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank M. Nussenzweig for providing reagents for antibody cloning and expression. This work was supported by the European Research Council (grant no. 250348 IMMUNExplore and 670955 BROADImmune), the Swiss National Science Foundation (grant no. 160279), the Swiss Vaccine Research Institute and the Wellcome Trust (grant no. 084535, 077092, 084538, 084113/Z/07/Z, 084378/Z/07/A, 092741 and 099811). A.L. is supported by the Helmut Horten Foundation. This paper is published with the permission of the Director of Kenya Medical Research Institute (KEMRI).

**Author Contributions** J.T. performed all experiments involving *P. falciparum*; K.P. characterized genomic DNA; L.P. produced mutant antibodies; J.T., K.P. and L.P. analysed the data and wrote the manuscript; A.A. and C.M.T. performed initial parasite work; M.F. performed bioinformatics analysis; R.G. analysed MS data; D.J. and C.S.F. performed cell sorting and antibody isolation; F.M.N., J.W. and Ph.B. provided cohort samples; B.F.-R. and So.B. produced antibodies; Si.B. performed immunoprecipitation experiments; K.M., V.T., D.C. and F.S. provided supervision; A.L. and Pe.B. provided overall supervision and wrote the manuscript.

**Author Information** The VH and VL sequences of the antibodies have been deposited in GenBank under accession numbers from KU058438 to KU058491 (Supplementary Table 1). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.L. ([lanzavecchia@irb.usi.ch](mailto:lanzavecchia@irb.usi.ch)) or Pe.B. ([pb642@cam.ac.uk](mailto:pb642@cam.ac.uk)).

## METHODS

**Parasite culture and selection.** The *Plasmodium falciparum* clone 3D7 and nine laboratory-adapted parasite isolates from severe and non-severe malaria patients in Kilifi, Kenya (sampled between 2009 and 2010), were cultured *in vitro* according to standard procedures<sup>20</sup> and cryopreserved at the late trophozoite stage for use in subsequent assays. To select for MGD21-reactive infected erythrocytes (IEs), cultured IEs were incubated with MGD21 for 20 min at room temperature, washed, and rotated with Protein G-coated magnetic beads (Life Technologies) for 30 min at room temperature. Following magnetic sorting, enriched (MGD21<sup>+</sup>) and depleted (MGD21<sup>-</sup>) fractions were returned to *in vitro* culture.

**Patients.** Donors C and D are 29 and 38 years old, respectively, and are lifelong residents of an area with moderate malaria transmission intensity (that is, with an entomological inoculation rate of 21.7 infective bites per person per year)<sup>21</sup>. Adults in this area are clinically immune from febrile malaria, having acquired immunity during childhood. The two donors were *P. falciparum*-negative during sample collection. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

**Triple mixed agglutination assay.** Following informed consent, plasma samples were taken from 2007 to 2014 from 557 adults living in a malaria-endemic region within Kilifi County on the coast of Kenya. The study was approved by the Kenya Medical Research Institute Ethics Review Committee and the Oxford Tropical Research Ethics Committee. IEs from three parasite isolates were separately stained with 10 µg ml<sup>-1</sup> DAPI, 200 µg ml<sup>-1</sup> ethidium bromide or 6.7 × SYBR Green I for 1 h at room temperature. The stained parasites were washed five times, mixed in equal proportions, and diluted to a 5% haematocrit in incomplete RPMI medium. Ten microlitres of the parasite mixture was rotated with 2.5 µl of adult plasma for 1.5 h at room temperature, and agglutinates formed were examined by fluorescence microscopy. In the primary screen, pools of five adult plasma were tested against six Kenyan isolates (in two separate reactions). Pools that formed mixed-colour agglutinates were identified and individual plasma within these pools were tested against nine isolates using the same assay. A single isolate (10668) was not detected in mixed agglutinates formed by any of the plasma and was therefore excluded from the study. Two adults (donors C and D) with plasma that formed mixed agglutinates with eight parasite isolates were selected for monoclonal antibody isolation and, following further informed consent, an additional blood sample was taken from each individual in February 2014.

**B-cell immortalization and isolation of monoclonal antibodies.** IgG<sup>+</sup> memory B cells were isolated from cryopreserved peripheral blood mononuclear cells (PBMCs) by magnetic cell sorting with mouse anti-CD19-PECy7 antibodies (BD Pharmingen, catalogue no. 341113) and mouse anti-PE microbeads (Miltenyi Biotec, catalogue no. 130-048-081), followed by flow cytometry sorting for IgG<sup>+</sup> IgM<sup>-</sup> IgD<sup>-</sup> cells. The B cells were immortalized with Epstein-Barr virus (EBV) in the presence of CpG-DNA (2.5 µg ml<sup>-1</sup>) and irradiated feeder cells as described previously<sup>5</sup>. Two weeks post-immortalization, culture supernatants were tested for the ability to stain IEs from eight parasite isolates by flow cytometry. Cryopreserved IEs were thawed, stained with 10 × SYBR Green I, and incubated with the B-cell supernatants for 1 h at 4 °C. Antibody binding was detected using 2.5 µg ml<sup>-1</sup> of goat Alexa Fluor 647-conjugated anti-human IgG (Jackson ImmunoResearch, catalogue no. 109-606-170). Reactivity was calculated based on the percentage of late-stage parasites (high SYBR Green) recognized by each antibody.

**Sequence analysis of antibody cDNA and genomic DNA.** cDNA was synthesized from selected B-cell cultures and both heavy chain and light chain variable regions (VH and VL) were sequenced as previously described<sup>22</sup>. The usage of VH and VL genes and the number of somatic mutations were determined by analysing the homology of VH and VL sequences of monoclonal antibodies to known human V, D and J genes in the IMGT database<sup>23</sup>. Genomic DNA was isolated from two B-cell lines of donor C and one B-cell line of donor D with a commercial kit (QIAGEN), and antibody-encoding sequences were amplified and sequenced with primers specific for the V and J regions of the given antibody. Sequences were aligned with ClustalW2 (ref. 24). Potential cryptic RSS sites were identified using the RSSsite web server<sup>25</sup>. To determine the heterozygosity of *LAIR1* on chromosome 19, the following primers were used to perform PCRs on genomic DNA: LAIR1\_INTR\_FW1, GGCGGTGGGCACTCAGGTTTC; LAIR1\_INTR\_REV1, CACAGGCAGTCACCGGGTCTAGG; LAIR1\_INTR\_FW2, GGATGCACCATGTCACCCAGTCTCTGG. Genomic DNA isolated from PHA- and IL-2-stimulated T cells from donor C was used as a control for sequence analysis.

**Immunoglobulin lineage and genealogy analysis.** Unmutated common ancestor (UCA) sequences of the VL region were inferred with Antigen Receptor Probabilistic Parser (ARPP) UA Inference software, as previously described<sup>26</sup>. UCA sequences of the VH region were constructed using IMGT/V-QUEST<sup>23</sup> and the genomic insert sequences. Nucleotide sequences of the mutated antibodies and the UCA were aligned using ClustalW2 (ref. 24), and phylogenetic trees were

generated with the DNA Maximum Likelihood program (Dnaml) of the PHYLIP package, version 3.69 (refs 27, 28).

**Production of recombinant antibodies, antibody variants and fusion proteins.** Antibody heavy and light chains were cloned into human IgG1, Igκ and Igλ expression vectors<sup>22</sup> and expressed by transient transfection of Expi293F Cells (ThermoFisher Scientific) using polyethylenimine (PEI). Cell lines were routinely tested for mycoplasma contamination. The antibodies were affinity purified by protein A chromatography (GE Healthcare). Variants of the MGD21 antibody were produced by (1) exchanging V<sub>H</sub>, D<sub>H</sub>, J<sub>H</sub> elements or the light chain with the corresponding sequences of an irrelevant antibody (FI499, reactive to influenza virus<sup>28</sup>), (2) deleting selected segments, or (3) reverting somatic mutations to the germline configuration with reference to the IMGT database and the original *LAIR1* genomic sequence (NCBI reference sequence NC\_018930.2). In addition, LAIR1-Fc fusion proteins were produced recombinantly by cloning the mutated or unmutated *LAIR1* fragment into a plasmid designed for expression of human IgG1 fusion proteins (pINFUSE-hIgG1-Fc2, Invivogen). On the basis of an alignment of the most potent LAIR1-containing antibodies with the unmutated LAIR1 sequence, five key residues that could contribute to gain of binding to IEs and loss of binding to collagen were identified and added alone or in various combinations to the unmutated LAIR1-Fc fusion protein. The MGD21 constructs and LAIR1 domain mutants were tested for staining of 3D7 IEs that were enriched for MGD21 recognition (3D7-MGD21<sup>+</sup>). For the LAIR1 domain mutants, MFI values at 1 µg ml<sup>-1</sup> antibody concentration were calculated by interpolation of binding curves fitted to a linear regression model (Graphpad Prism 6).

**ELISA.** Total IgGs were quantified using 96-well MaxiSorp plates (Nunc) coated with goat anti-human IgG (SouthernBiotech, catalogue no. 2040-01) using Certified Reference Material 470 (ERMs-DA470, Sigma-Aldrich) as a standard. To test binding to human collagen type I, ELISA plates were coated with 5 µg ml<sup>-1</sup> of type I recombinant human collagen (Millipore, catalogue no. CC050), blocked with 1% bovine serum albumin (BSA) and incubated with titrated antibodies, followed by AP-conjugated goat anti-human IgG, Fcγ fragment specific (Jackson ImmunoResearch, catalogue no. 109-056-098). Plates were then washed, substrate (p-NPP, Sigma) was added and plates were read at 405 nm.

**Immunoprecipitation and LC-MS.** Erythrocyte ghosts were prepared by hypotonic lysis with 1 × PBS diluted 15-fold in water, and ghost membranes were dissolved in a reducing lysis buffer containing 2% SDS, 10 mM dithiothreitol (DTT), 10 mM HEPES pH 8, sonicated and boiled. Solubilized proteins were alkylated with iodoacetamide (final concentration 55 µM) for 30 min at room temperature and precipitated with 80% acetone overnight at 4 °C. The precipitates were resuspended in urea and digested with trypsin. For immunoprecipitation experiments, IEs were sonicated and dissolved in 7.2 M urea in RIPA buffer (1% Triton X-100, 0.1% SDS, 0.5% sodium deoxycholate in HBS pH 7.4). The samples were centrifuged and supernatants were diluted 6.7-fold with RIPA buffer containing a protease inhibitor cocktail (Sigma-Aldrich) and incubated with 10 µg of MGD21 or BKC3 overnight at 4 °C. Next, Protein G-Sepharose beads (GE Healthcare) were added and samples were incubated for 1 h at 4 °C. The beads were washed four times and immunoprecipitates were digested directly on the beads with trypsin. After trypsin digestion, peptides were analysed on a Q-Exactive instrument at the Functional Genomics Center in Zurich. Raw files were analysed using the MaxQuant software<sup>29,30</sup> and MS/MS spectra were searched against the human and *P. falciparum* 3D7 UniProt FASTA databases (UP000005640 and UP000001450). Peptide identifications were matched across several replicates. Subsequent data analysis was performed in the R statistical computing environment. Missing values were imputed with a normal distribution around an LFQ value of 21. Statistical significance was evaluated by Welch tests.

**Western blots.** Ghosts and immunoprecipitates were dissolved in 2 × SDS sample buffer (Bio-Rad) and run on a 12% polyacrylamide gel under non-reducing conditions. The proteins on the gel were transferred onto a PVDF membrane, which was blocked with 5% milk in TBS with 0.1% Tween (TBST) for 1 h at room temperature. The membrane was incubated with 5 µg ml<sup>-1</sup> MGD21 overnight at 4 °C, washed with TBST, and developed with horseradish peroxidase (HRP)-conjugated sheep anti-human IgG (GE Healthcare, catalogue no. NA933) used in combination with a chemiluminescent substrate.

**Expression of RIFINs.** Genes encoding the A-RIFINs PF3D7\_1400600, PF3D7\_1040300, PF3D7\_0100400, PF3D7\_0100200 and PF3D7\_1100500 were produced by gene synthesis (Genscript) and cloned into the pDisplay vector (Invitrogen), which contains a haemagglutinin (HA) tag, as previously described<sup>3</sup>. RIFIN chimaeras containing the constant region of PF3D7\_1400600 (residues 38–146) and the variable region of PF3D7\_0100200 (residues 151–288) (PF3D7\_1400600c\_0100200v), or containing the constant region of PF3D7\_0100200 (residues 42–150) and the variable region of PF3D7\_1400600 (residues 147–325) (PF3D7\_0100200c\_1400600v), were generated. The pDisplay



constructs were transiently transfected into CHOK1-SV cells (GS-System, Lonza) using PEI. Cell lines were routinely tested for mycoplasma contamination. Briefly, 1 day before transfection, CHOK1-SV cells were seeded at  $0.5 \times 10^6$  cells  $\text{ml}^{-1}$  in 30 ml CD-CHO medium (Invitrogen) supplemented with 2 mM L-glutamine in 125 ml Erlenmeyer flasks (Corning). On the day of transfection, 20  $\mu\text{g}$  DNA was diluted in OPTI-PRO SFM Medium (Invitrogen) and mixed with 200  $\mu\text{g}$  PEI for 20 min at room temperature. The DNA-PEI complexes were added to the cells, which were cultured in a  $\text{CO}_2$  shaker incubator at  $37^\circ\text{C}$ , 135 r.p.m. After 72 h, the expression of RIFINs and their recognition by the LAIR1-containing antibodies were tested by flow cytometry. Briefly, 5  $\mu\text{g}$   $\text{ml}^{-1}$  of rabbit anti-HA tag and 2  $\mu\text{g}$   $\text{ml}^{-1}$  of MGC or MGD antibodies were added to the RIFIN-transfected cells. Antibody binding was detected by 5  $\mu\text{g}$   $\text{ml}^{-1}$  of Alexa Fluor 488-conjugated goat anti-rabbit IgG (Life Technologies, catalogue no. A11034) and 2.5  $\mu\text{g}$   $\text{ml}^{-1}$  of Alexa Fluor 647-conjugated goat anti-human IgG (Jackson ImmunoResearch, catalogue no. 109-606-170). Dead cells were excluded by staining with 7-AAD (BD Biosciences).

**Inhibition of parasite growth.** 3D7-MGD21<sup>+</sup> (5% parasitaemia, ring stage) was cultured with various concentrations of MGD21 or BKC3 for 2 days. After 2 days, 10 $\times$  SYBR Green I was added to aliquots of each culture and parasitaemia was quantified by flow cytometry. The remaining parasites in each culture were washed to remove the antibodies and incubated for 1 day to allow the parasites to reach the late trophozoite/schizont stage. MGD21 recognition of these cultures was detected using 2.5  $\mu\text{g}$   $\text{ml}^{-1}$  of Alexa Fluor 647-conjugated goat anti-human IgG (Jackson ImmunoResearch, catalogue no. 109-606-170).

**Inhibition of rosetting.** 9605-MGD21<sup>+</sup> IEs at the late trophozoite/schizont stage were purified from uninfected erythrocytes and ring-stage parasites using a magnetic column (Miltenyi Biotec) and were resuspended in culture medium with 10% human serum. The purified IEs were incubated with 10  $\mu\text{g}$   $\text{ml}^{-1}$  of MGD21 or BKC3 for 1 h at  $4^\circ\text{C}$ , mixed with O<sup>+</sup> erythrocytes or A<sup>+</sup> erythrocytes in a 1:20 ratio, and incubated for 30 min at room temperature to allow rosetting to occur. The IEs were stained with 10 $\times$  SYBR Green I, and the number of rosettes formed by at least 200 IEs was counted by fluorescence microscopy to calculate the rosetting rate.

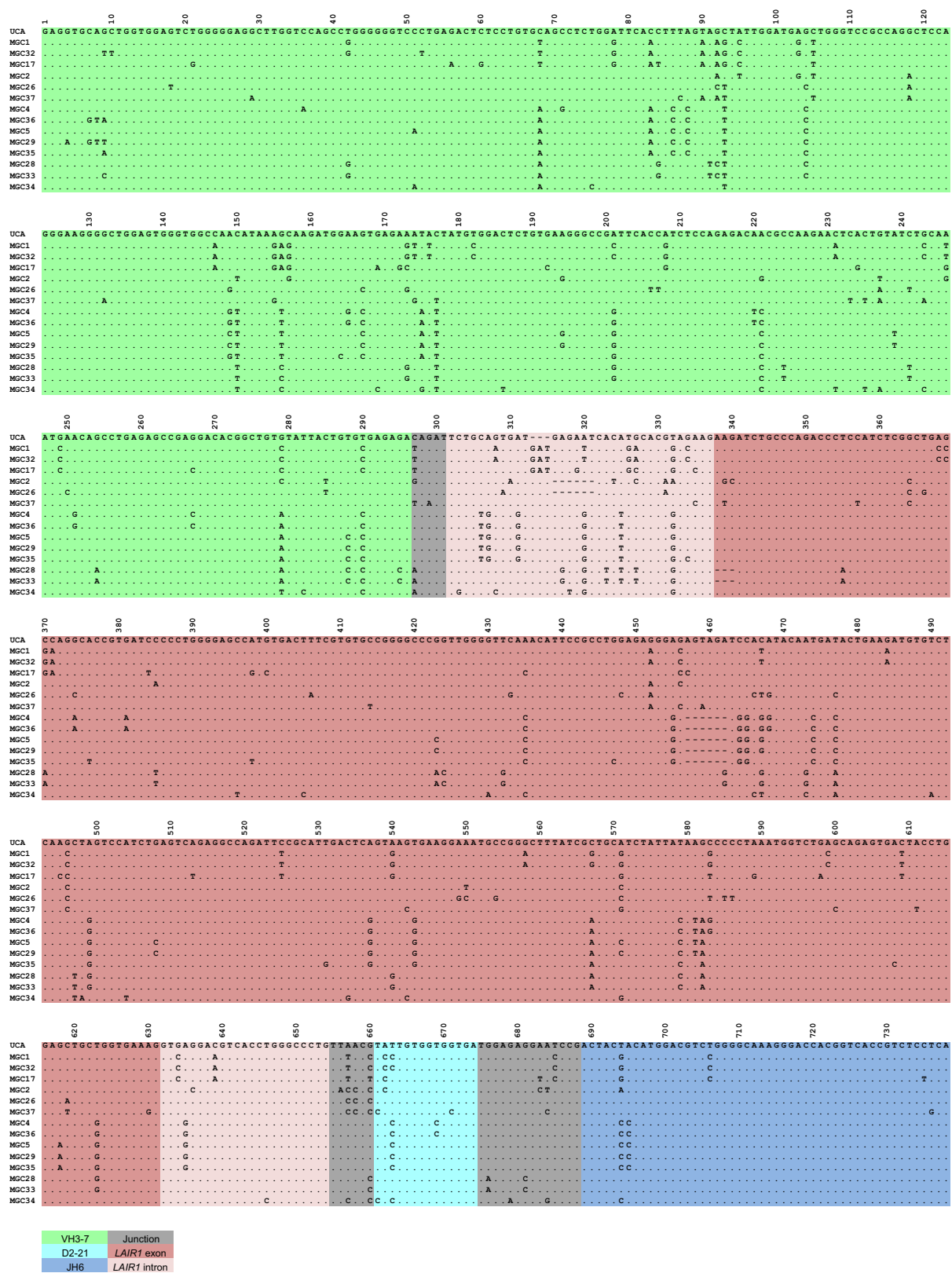
**Agglutination with monoclonal antibodies.** 3D7-MGD21<sup>+</sup> and 11019-MGD21<sup>+</sup> IEs (4–5% parasitaemia) were diluted to a 3% haematocrit in a 5 $\times$  SYBR Green I solution containing 5  $\mu\text{g}$   $\text{ml}^{-1}$  of the test monoclonal antibody. Each sample was rotated for 1 h at room temperature and subsequently examined by fluorescence microscopy.

**Opsonic phagocytosis by monocytes.** IEs were stained with 10  $\mu\text{g}$   $\text{ml}^{-1}$  DAPI for 30 min at room temperature, washed four times and run on a magnetic column (Miltenyi Biotec) to purify late-stage parasites. The purified parasites were opsonized with serially diluted antibodies for 1 h at  $4^\circ\text{C}$ . Monocytes were isolated from fresh PBMCs of healthy donors using mouse anti-CD14 microbeads (Miltenyi Biotec, catalogue no. 130-050-201) and mixed with the opsonized parasites in a 1:2 ratio for 1 h at  $37^\circ\text{C}$ . Extracellularly bound, non-internalized IEs were lysed by treatment with red blood cell lysis solution (Miltenyi Biotec) for 10 min at room temperature. The cells were stained with mouse anti-CD14-PECy5 (Beckman Coulter, catalogue no. A07765) and analysed by flow cytometry. The mean fluorescence intensity (MFI) of DAPI in CD14<sup>+</sup> cells was used as a measure of phagocytosis of IEs by monocytes.

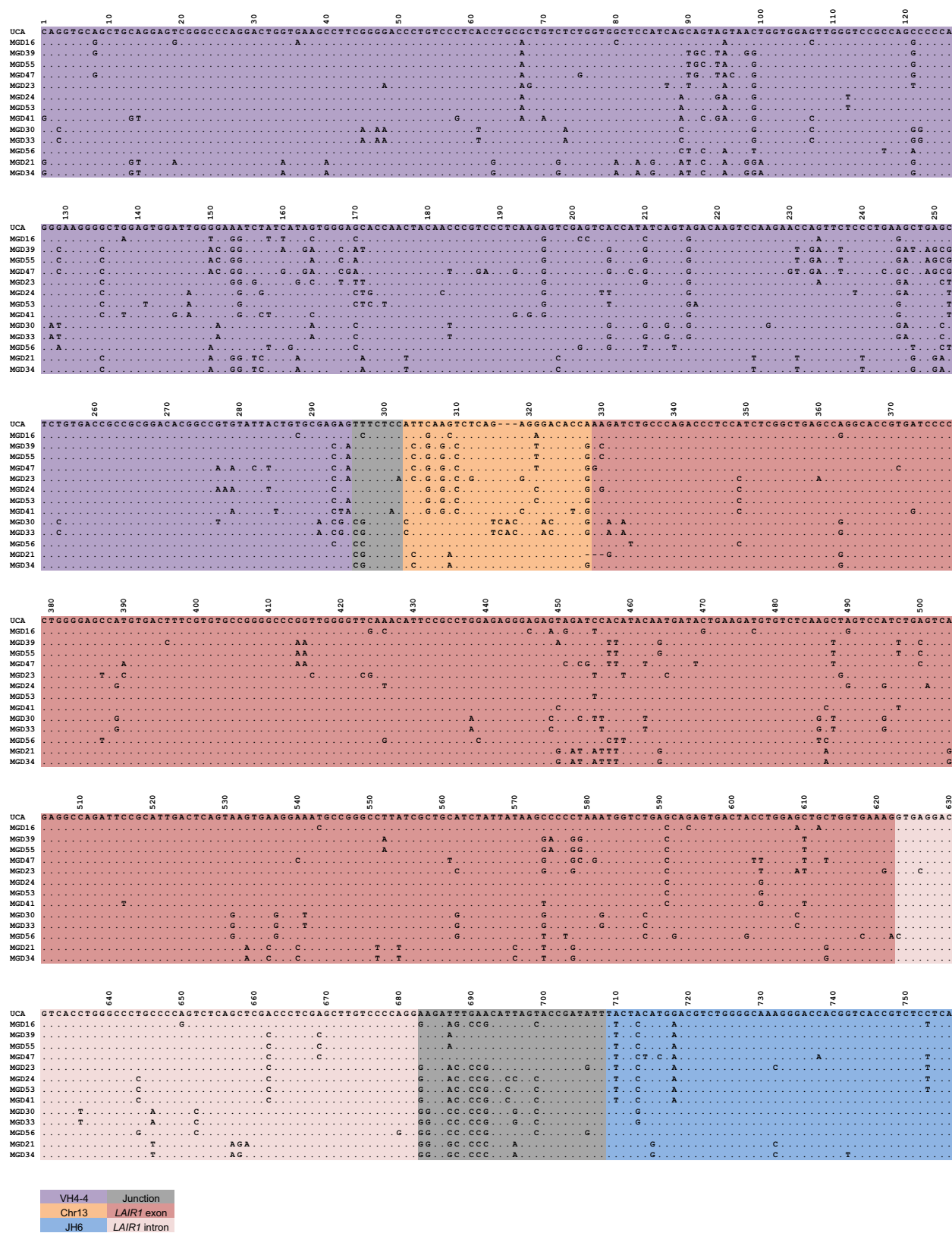
**Statistics.** The Wilcoxon signed-rank test was used for statistical comparisons of pairs of data groups in rosetting experiments. No statistical methods were used to predetermine sample size.

20. Trager, W. & Jensen, J. B. Human malaria parasites in continuous culture. *Science* **193**, 673–675 (1976).
21. Midega, J. T. *et al.* Wind direction and proximity to larval sites determines malaria risk in Kilifi District in Kenya. *Nat. Commun.* **3**, 674 (2012).
22. Tiller, T. *et al.* Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning. *J. Immunol. Methods* **329**, 112–124 (2008).
23. Lefranc, M.-P. *et al.* IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res.* **37**, D1006–D1012 (2009).
24. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
25. Merelli, I. *et al.* RSSsite: a reference database and prediction tool for the identification of cryptic recombination signal sequences in human and murine genomes. *Nucleic Acids Res.* **38**, W262–W267 (2010).
26. Kepler, T. B. Reconstructing a B-cell clonal lineage. I. Statistical inference of unobserved ancestors. *F1000 Res.* **2**, 103 (2013).
27. Liao, H.-X. *et al.* Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* **496**, 469–476 (2013).
28. Pappas, L. *et al.* Rapid development of broadly influenza neutralizing antibodies through redundant mutations. *Nature* **516**, 418–422 (2014).
29. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnol.* **26**, 1367–1372 (2008).
30. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).

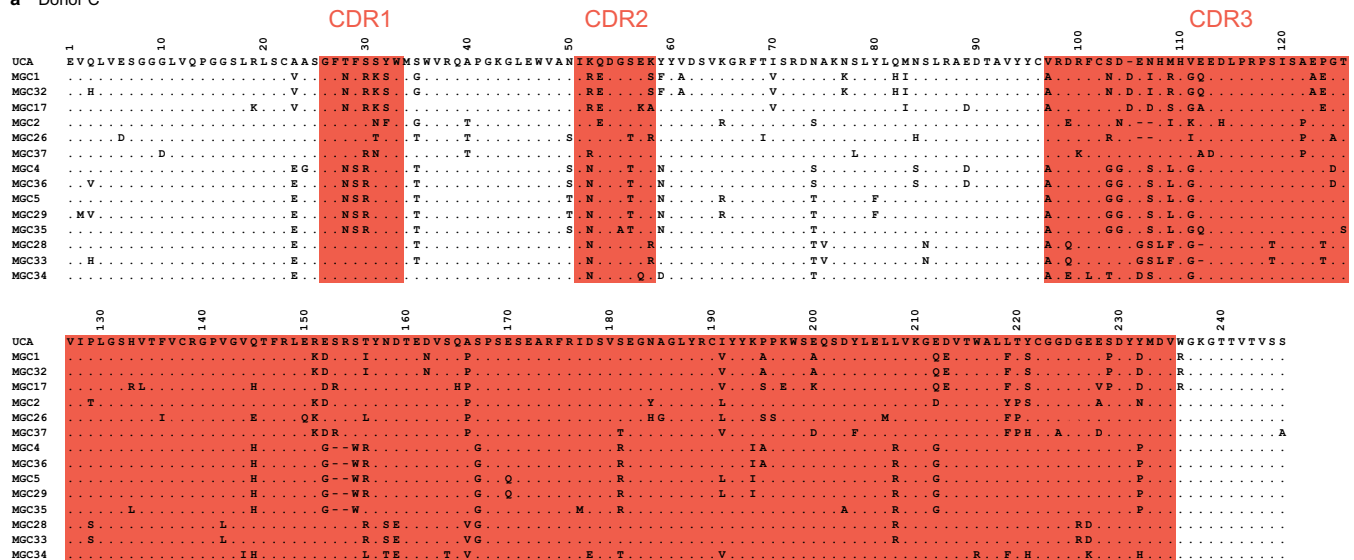




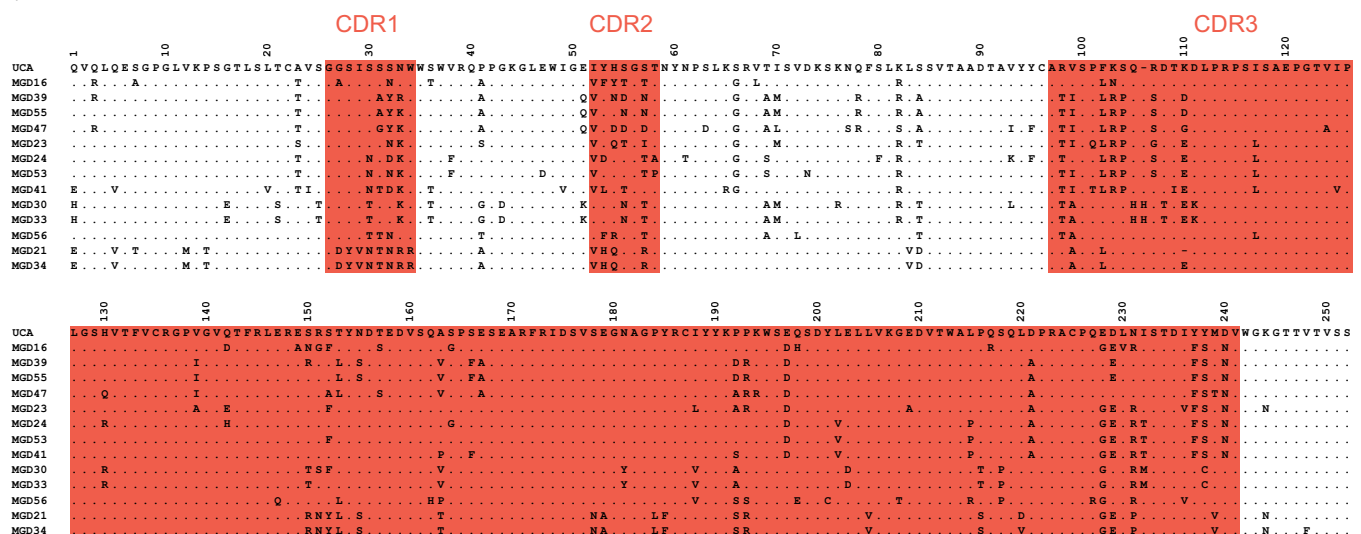
**Extended Data Figure 1 | Nucleotide sequence alignments of VH regions of antibodies isolated from donor C.** Dots indicate positions where the nucleotide of a mature antibody is identical to that of the UCA.



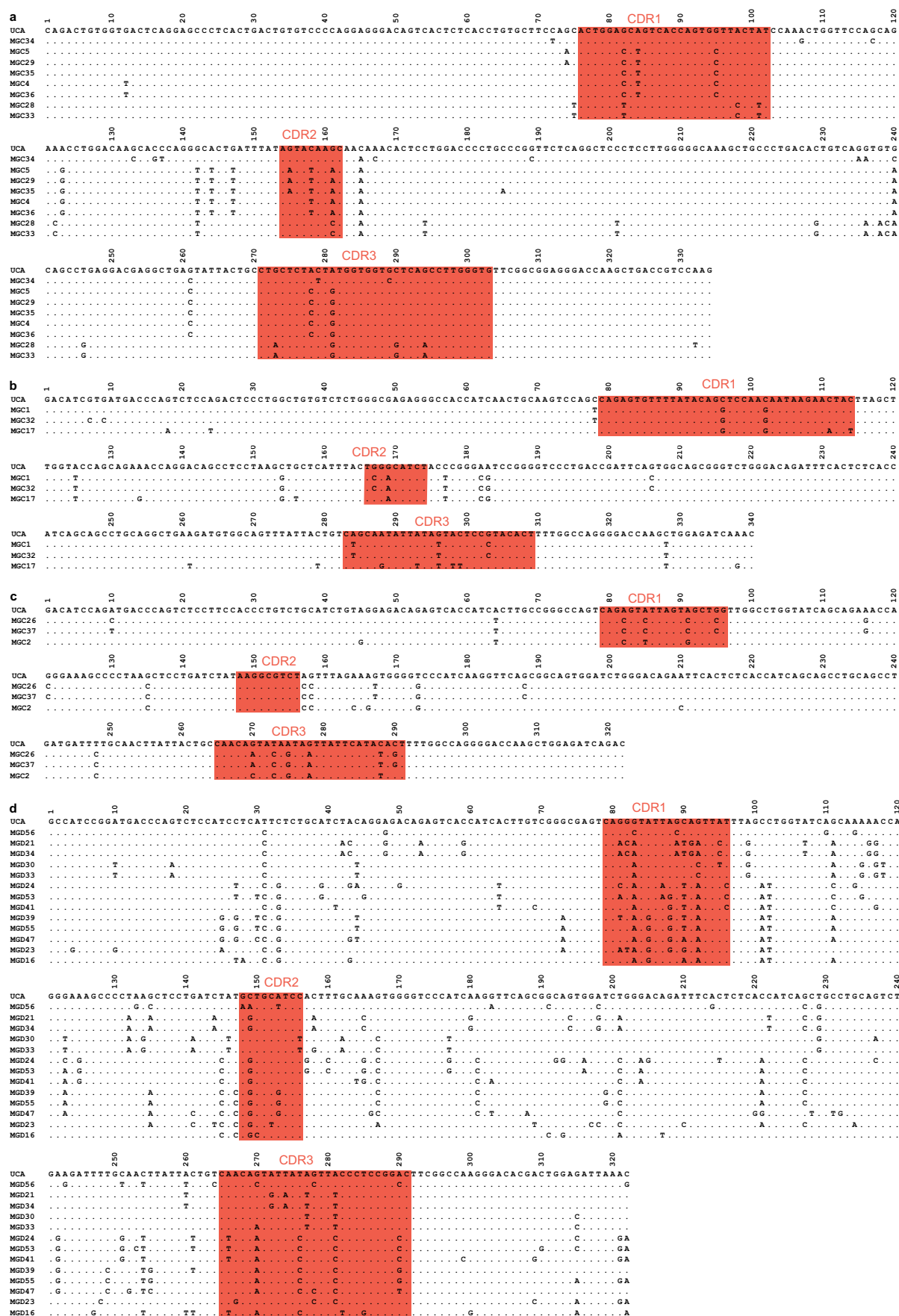
## a Donor C



## b Donor D



**Extended Data Figure 3 | Protein sequence alignments of VH regions of antibodies isolated from donors C and D. a, Donor C. b, Donor D.** Putative complementarity-determining regions (CDRs) are highlighted in red. Dots indicate positions where the amino acid of a mature antibody is identical to that of the UCA.

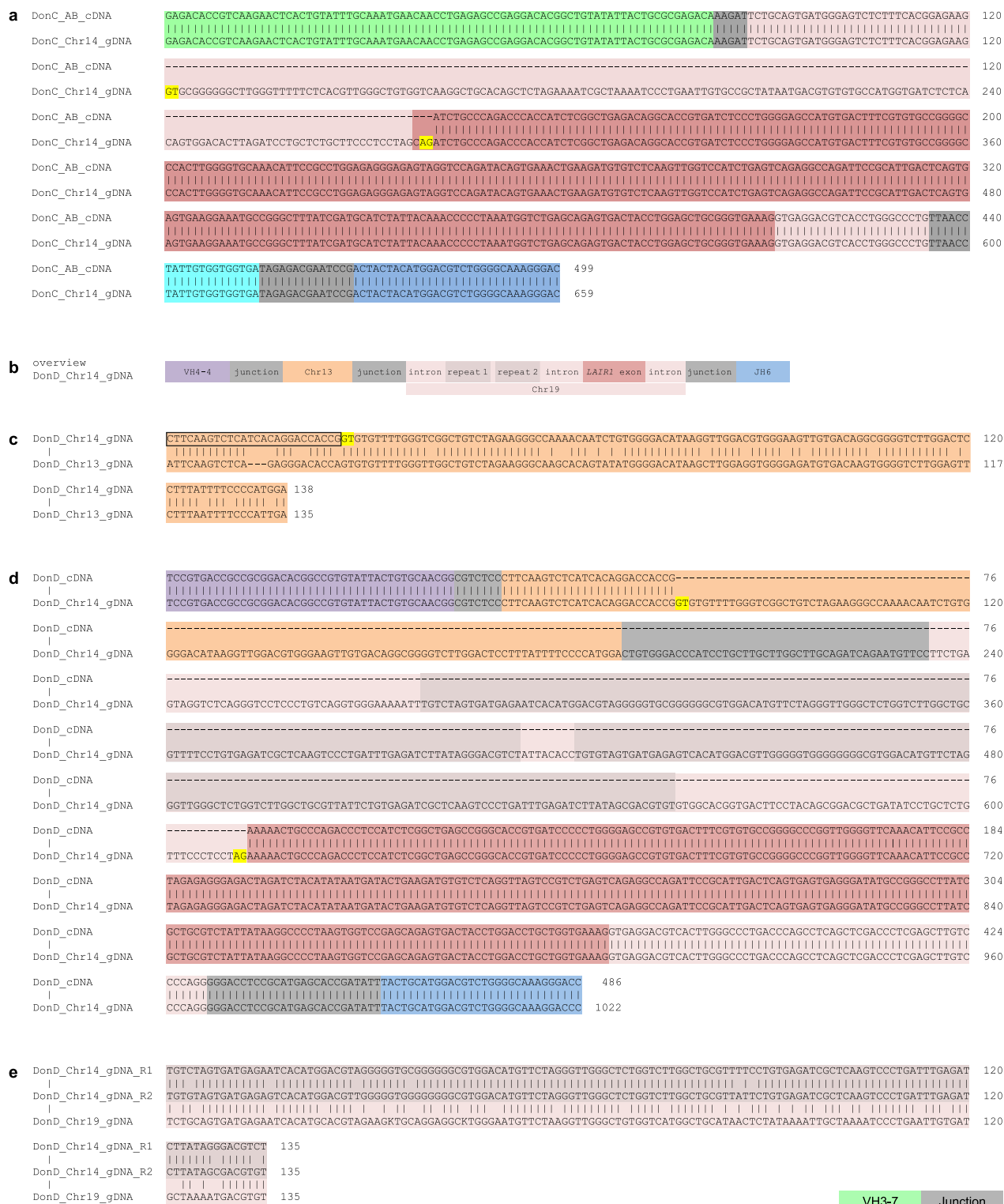


**Extended Data Figure 4 | Nucleotide sequence alignments of VL regions of antibodies isolated from donors C and D. a–d, Antibodies from donor C use VL7-43/JL3 (a), VK4-1/JK2 (b), or VK1-5/JK2 (c), while antibodies**

**from donor D use VK1-8/JK5 (d). Complementarity-determining regions (CDRs) are highlighted in red. Dots indicate positions where the nucleotide of a mature antibody is identical to that of the UCA.**







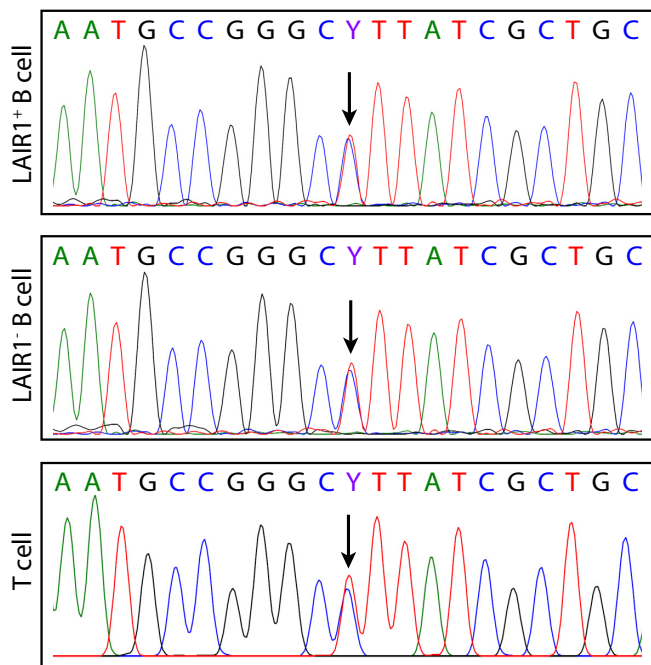
**Extended Data Figure 6 | Genomic DNA analysis of LAIR1-containing antibodies of donor C and donor D.** **a**, The sequence alignment of genomic DNA (gDNA) and cDNA of a LAIR1-containing antibody (AB) from donor C (DonC) reveals a 507 bp *LAIR1* insert in chromosome 14 (Chr14) and the removal of a 160 bp fragment by RNA splicing. Splice donor and acceptor sites are highlighted in yellow. **b**, Schematic overview of the genomic organization of a LAIR1-containing antibody from donor D, not to scale. **c**, Alignment of a region of antibody-encoding DNA (chromosome 14) with

the corresponding region of chromosome 13 from gDNA. The sequence maintained in the mature antibody mRNA is boxed and the splice donor site is highlighted in yellow. **d**, Alignment of gDNA and cDNA reveals that a part of the chromosome 13 region and the entire inserted 5' *LAIR1* intron are removed by RNA splicing. Splice donor and acceptor sites are highlighted in yellow. **e**, Alignment of the two repeated elements found in the inserted *LAIR1* intron in chromosome 14 with the corresponding sequence in chromosome 19. The repeats are named R1 and R2, and K = G/A.

	(Score)	12-cryptic-RSS	23-cryptic-RSS-antiparallel	(Score)
MGC_Ch19	(-57.29)	TCTG <u>CAGTGATGAGAATCACATGCACGTAGAA</u> .....	<u>GAGCTGCTGGTGAAAGGTGAGGACGTCACCTGGGCCCTG</u>	(-79.68)
MGD_Ch19	(-62.84)	TTGTGAG <u>CAAGTCTCAGGGTCCTCACTGTCAACTG</u> .....	<u>CTGGGCCCTGCCCCAGTCTCAGCTCGACCCCTCGAGCTTG</u> TCCCCAGG	(-77.42)
MGD_Ch13	(-64.12)	ATT <u>CAAGTCTCAGAGGGACACCAGTGTGTTT</u> .....	<u>TGACAAGTGGGGTCTTGAGTTCTTTAATTTTCCCATTGA</u>	(-75.63)

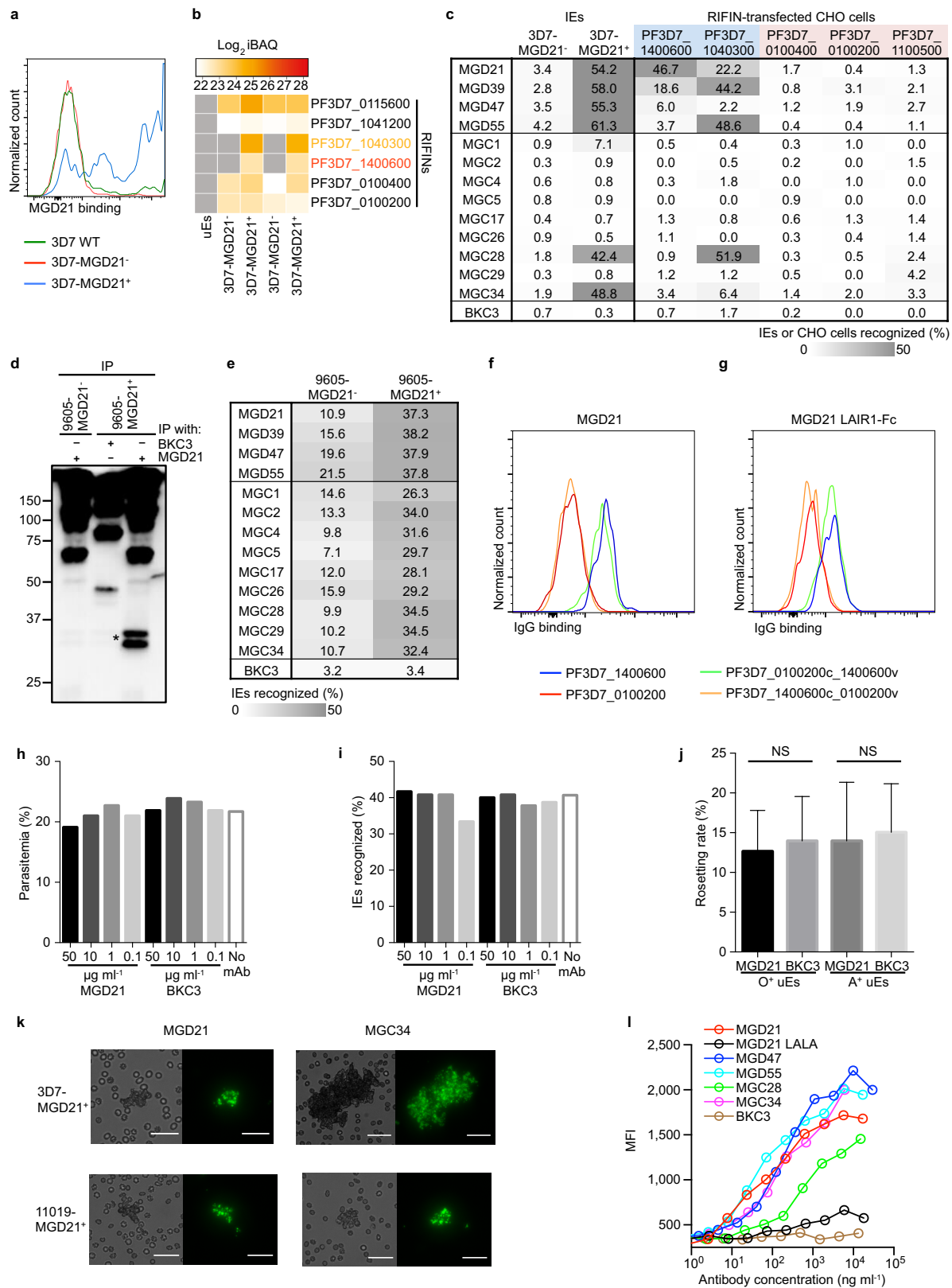
**Extended Data Figure 7 | *LAIR1* and chromosome 13 inserts are flanked by 12/23 cryptic RSS sites.** The regions on chromosome 19 (Chr19) and chromosome 13 (Chr13) of donor-derived gDNA corresponding to the ends of the inserts were sequenced and RSS sites were identified using

the RSSsite web server. The sequences shown begin from the ends of the inserts. Cryptic RSS sites are highlighted in grey, with complementary ends underlined and prediction scores shown in parentheses.



**Extended Data Figure 8 | Both *LAIR1* alleles on chromosome 19 are intact in B cells producing LAIR1 antibodies.** Heterozygosity of the chromosome 19 *LAIR1* exon in cells from donor C showing that both *LAIR1* alleles are intact in B cells producing LAIR1-containing antibodies. Displayed are the chromatograms obtained for B-cell clones with or without a LAIR1 insertion (LAIR1<sup>+</sup> or LAIR1<sup>-</sup> B cell) and for polyclonal T cells. Y = C/T

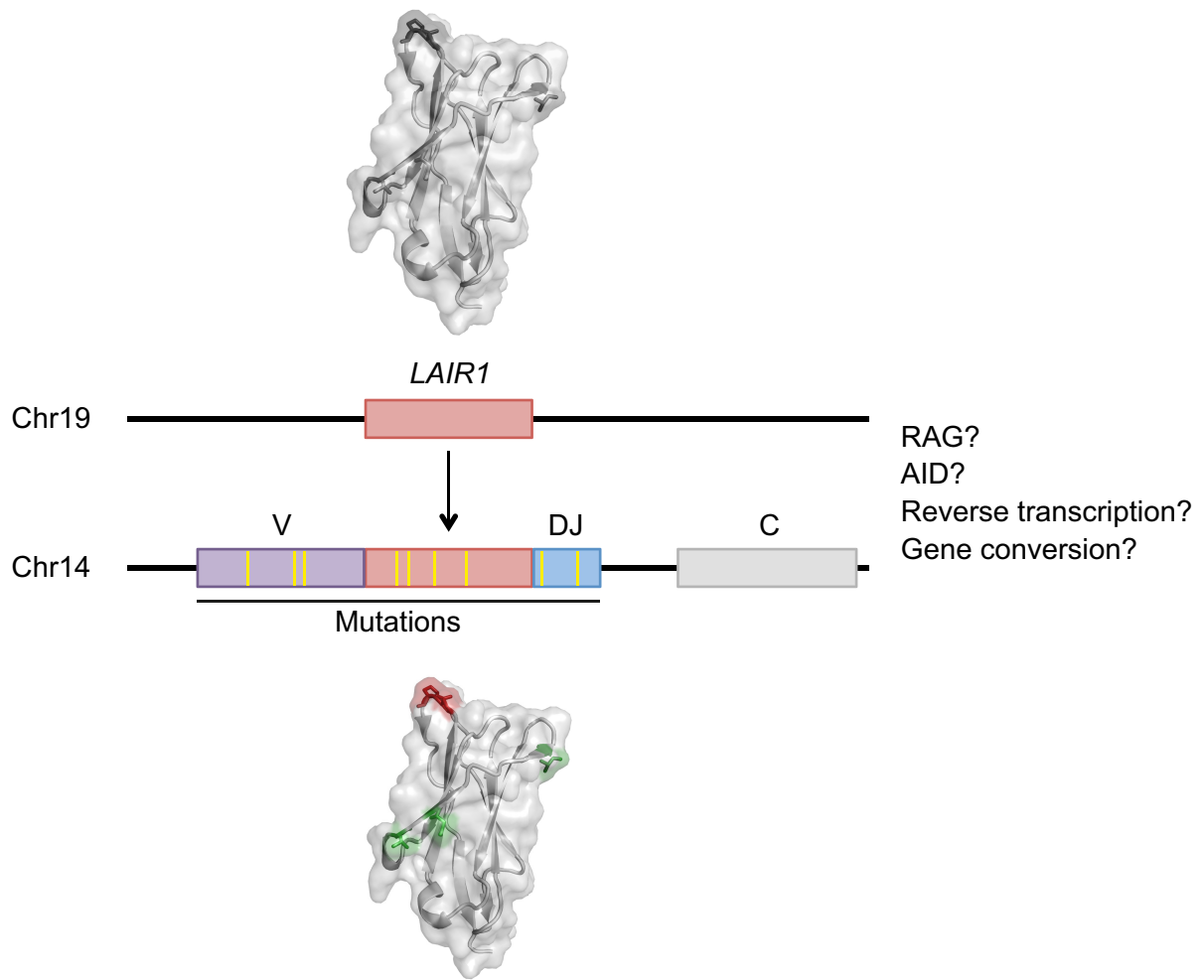




Extended Data Figure 9 | See next page for figure caption.

**Extended Data Figure 9 | Reactivity and functional assays of MGC and MGD antibodies.** **a**, MGD21 staining of 3D7 IEs that were enriched or depleted of MGD21 reactivity (representative of  $n = 3$  independent experiments). WT, wild type. **b**, Heat map from LC-MS analysis showing RIFIN expression levels (calculated as intensity-based absolute quantification (iBAQ) scores) in erythrocyte ghosts prepared from 3D7-MGD21<sup>+</sup> and 3D7-MGD21<sup>-</sup> IEs (two experiments shown). Grey boxes indicate that expression levels are below the detection limit. **c**, Shown is the percentage of IEs (representative of  $n = 2$  independent experiments) or of transfected CHO cells ( $n = 1$ ) stained by the antibodies. RIFINs that were enriched in 3D7-MGD21<sup>+</sup> ghosts are highlighted blue, while RIFINs that were similarly expressed or not detected in 3D7-MGD21<sup>-</sup> and 3D7-MGD21<sup>+</sup> ghosts are shown in red. BKC3 is a negative control antibody. **d**, Western blot showing MGD21 binding to immunoprecipitates (IP) prepared from 9605-MGD21<sup>-</sup> and 9605-MGD21<sup>+</sup> IEs (representative of  $n = 2$  independent experiments). Specific bands are marked with an asterisk. Anti-human IgG was used as the secondary antibody, resulting in detection of antibodies used for immunoprecipitation alongside antigens of interest. For gel source data, see Supplementary Fig. 1. Numbers on left indicate kDa **e**, Percentage of 9605-MGD21<sup>-</sup> and 9605-MGD21<sup>+</sup> IEs recognized by representative MGC and MGD antibodies (representative of  $n = 2$  independent experiments). **f**, Binding of MGD21 to CHO cells

transfected with RIFINs (PF3D7\_1400600 and PF3D7\_0100200), a RIFIN chimera containing the constant region of PF3D7\_0100200 and the variable region of PF3D7\_1400600 (PF3D7\_0100200c\_1400600v), or the inverse chimera (PF3D7\_1400600c\_0100200v) ( $n = 1$ ). **g**, Binding of an Fc fusion protein containing the LAIR1 domain of MGD21 to CHO cells transfected with RIFINs or RIFIN chimaeras ( $n = 1$ ). **h**, Parasitaemia of 3D7-MGD21<sup>+</sup> *in vitro* culture after 2 days of incubation with various concentrations of MGD21 or an irrelevant antibody (BKC3) ( $n = 1$ ). mAb, monoclonal antibody. **i**, Percentage of 3D7-MGD21<sup>+</sup> IEs recognized by MGD21 after 2 days of incubation with various concentrations of MGD21 or BKC3. The antibodies were removed after 2 days (during the ring stage of the life cycle) and the parasites were allowed to grow for 24 h to the late trophozoite/schizont stage before detection with MGD21 ( $n = 1$ ). **j**, Rosetting of 9605-MGD21<sup>+</sup> IEs with blood group O<sup>+</sup> or A<sup>+</sup> uninfected erythrocytes (uEs) after incubation with MGD21 or BKC3. Shown is the mean  $\pm$  s.d. from  $n = 4$  independent experiments. Statistical significance was evaluated by the Wilcoxon signed-rank test ( $P > 0.1$  for both blood groups). NS, not significant. **k**, Agglutinates of 3D7-MGD21<sup>+</sup> or 11019-MGD21<sup>+</sup> IEs formed by MGD21 or MGC34. Scale bar, 25  $\mu$ m. **l**, Opsonic phagocytosis of 11019-MGD21<sup>+</sup> IEs by monocytes ( $n = 2$ ). The IEs were stained with DAPI, which was quantified in monocytes as a measure of phagocytosis.



**Extended Data Figure 10 | A schematic representation of interchromosomal *LAIR1* transposition.** Shown is the insertion of a fragment of *LAIR1* into the immunoglobulin heavy chain locus through

a mechanism still to be molecularly defined, followed by the acquisition of somatic mutations that increase binding to IEs and abolish binding to collagen. Chr, chromosome.

# Insulator dysfunction and oncogene activation in *IDH* mutant gliomas

William A. Flavahan<sup>1,2,3\*</sup>, Yotam Drier<sup>1,2,3\*</sup>, Brian B. Liao<sup>1,2,3</sup>, Shawn M. Gillespie<sup>1,2,3</sup>, Andrew S. Venteicher<sup>1,2,4</sup>, Anat O. Stemmer-Rachamimov<sup>1</sup>, Mario L. Suvà<sup>1,2</sup> & Bradley E. Bernstein<sup>1,2,3</sup>

**Gain-of-function *IDH* mutations are initiating events that define major clinical and prognostic classes of gliomas<sup>1,2</sup>. Mutant *IDH* protein produces a new onco-metabolite, 2-hydroxyglutarate, which interferes with iron-dependent hydroxylases, including the TET family of 5'-methylcytosine hydroxylases<sup>3-7</sup>. TET enzymes catalyse a key step in the removal of DNA methylation<sup>8,9</sup>. *IDH* mutant gliomas thus manifest a CpG island methylator phenotype (G-CIMP)<sup>10,11</sup>, although the functional importance of this altered epigenetic state remains unclear. Here we show that human *IDH* mutant gliomas exhibit hypermethylation at cohesin and CTCF-binding factor (CTCF)-binding sites, compromising binding of this methylation-sensitive insulator protein. Reduced CTCF binding is associated with loss of insulation between topological domains and aberrant gene activation. We specifically demonstrate that loss of CTCF at a domain boundary permits a constitutive enhancer to interact aberrantly with the receptor tyrosine kinase gene *PDGFRA*, a prominent glioma oncogene. Treatment of *IDH* mutant gliomaspheres with a demethylating agent partially restores insulator function and downregulates *PDGFRA*. Conversely, CRISPR-mediated disruption of the CTCF motif in *IDH* wild-type gliomaspheres upregulates *PDGFRA* and increases proliferation. Our study suggests that *IDH* mutations promote gliomagenesis by disrupting chromosomal topology and allowing aberrant regulatory interactions that induce oncogene expression.**

The human genome is organized into topological domains that represent discrete structural and regulatory units<sup>12</sup>. Such domains are evident in genome-wide contact maps generated by high-throughput chromatin conformation capture (HiC) techniques<sup>13</sup>, and have been termed 'topologically associated domains' or 'contact domains'<sup>14-16</sup>. Recent studies have strengthened the role of the CTCF insulator protein in creating chromatin loops and boundaries that partition such domains<sup>15</sup>. Genomic alterations that remove CTCF-associated boundaries allow aberrant enhancer-gene interactions and alter gene expression<sup>17</sup>.

Since CTCF binding is methylation-sensitive<sup>18,19</sup>, its localization might be altered by DNA hypermethylation in *IDH* mutant gliomas. We therefore used chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) to map CTCF binding genome-wide in 11 primary tumours and 4 glioma cell lines. Although CTCF binding patterns tend to be relatively stable, we detected highly overlapping subsets of CTCF sites that were lost in *IDH* mutants (Fig. 1a, b and Methods). Significantly more sites were commonly lost than gained (625 versus 300,  $P < 10^{-12}$ ). Whole-genome bisulfite sequencing data from The Cancer Genome Atlas (TCGA)<sup>10</sup> was used to assess the methylation status of 625 loci with reduced CTCF binding in mutant tumours. We found that these loci have higher GC content, and exhibit significantly higher levels of DNA methylation in *IDH* mutant gliomas relative to *IDH* wild type (Fig. 1c, d).

We considered that altered DNA methylation and CTCF binding might disrupt topological domain boundaries and gene insulation in *IDH* mutant tumours. We collated a set of constitutive domain boundaries based on kilobase (kb)-resolution HiC maps<sup>15</sup>. We then examined published RNA-seq expression data for 357 normal brain tissue samples<sup>20</sup>. Consistent with previous studies<sup>16</sup>, we found that genes in the same domain correlate across samples, but that genes separated by a boundary show lower correlation (Fig. 1e). We next incorporated expression data for 230 *IDH* mutant (218 *IDH1* mutant and 12 *IDH2* mutant) and 56 wild-type lower-grade gliomas, generated by TCGA<sup>2</sup>. Here again we found that the presence of an intervening boundary reduces correlation between neighbouring genes. We next scanned the genome for pairs of proximal genes separated by less than 180 kb (the average contact domain size<sup>15</sup>) that correlate much more strongly in *IDH* mutants than in wild-type gliomas (Fig. 1f and Methods). Remarkably, the resulting set is strongly enriched for gene pairs that cross domain boundaries (90% versus 69% expected at random;  $P < 10^{-4}$ ). Conversely, gene pairs that correlate less strongly in *IDH* mutants are more likely to reside in the same domain (52% versus 31% expected at random;  $P < 10^{-5}$ ). Notably, CTCF knockdown has been shown to increase cross-boundary interactions and decrease intra-domain interactions<sup>21</sup>. Thus, altered expression patterns in *IDH* mutant gliomas may reflect reduced CTCF binding and consequent disruption of domain boundaries and topologies.

We next sought to pinpoint specific boundaries that were disrupted by *IDH* mutation. For all pairs of genes separated by <1 megabase (Mb), we computed their correlation across mutant and wild-type *IDH* gliomas. We then scanned for loci in which cross-boundary gene pairs correlate more strongly in mutant tumours (false discovery rates (FDR) < 1%), while intra-domain gene pairs correlate less strongly (FDR < 1%). This analysis highlighted 203 domain boundaries (Fig. 2a, Supplementary Table 1 and Methods). The putatively disrupted boundaries exhibit higher DNA methylation and lower CTCF binding in *IDH* mutant compared with wild-type tumours (Extended Data Fig. 1). These data suggest that the methylator phenotype disrupts CTCF binding and domain boundaries, thereby affecting gene expression in *IDH* mutant gliomas.

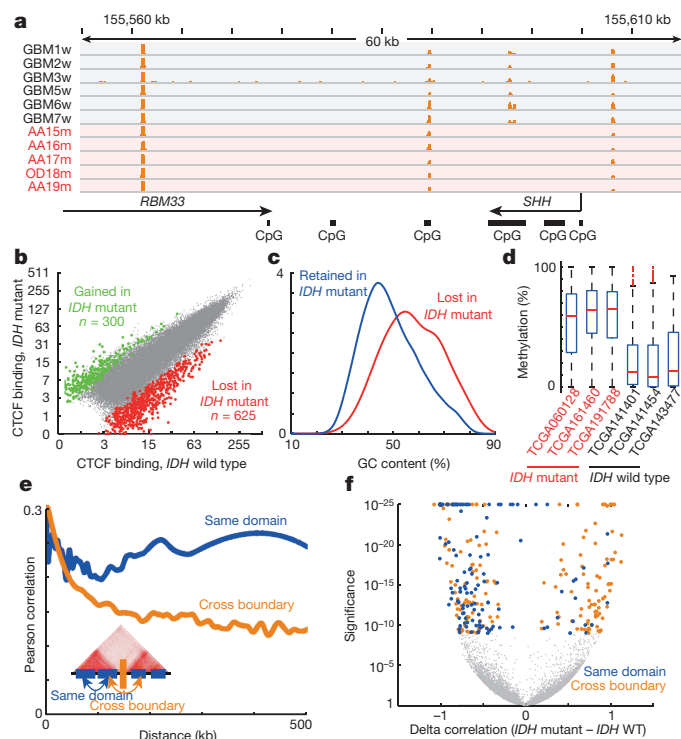
We hypothesized that altered domain topologies might contribute to gliomagenesis by activating oncogenes that are normally insulated by domain boundaries. We therefore scanned the domains adjacent to the disrupted boundaries for genes with higher expression in *IDH* mutant than in wild-type gliomas (Fig. 2a). Genes in top-scoring domains include *PDGFRA* ( $P < 10^{-21}$ ), an established glioma oncogene<sup>22</sup>, and other candidate regulators of gliomagenesis (Supplementary Table 1).

The identification of *PDGFRA* as a potential target of epigenetic deregulation in *IDH* mutants was of particular interest, given its prominence as a glioma oncogene and established roles for PDGFA signalling in the normal brain. Although *PDGFRA* is a frequent target of genomic

<sup>1</sup>Department of Pathology and Center for Cancer Research, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA. <sup>2</sup>Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. <sup>3</sup>Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA. <sup>4</sup>Department of Neurosurgery, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA.

\*These authors contributed equally to this work.

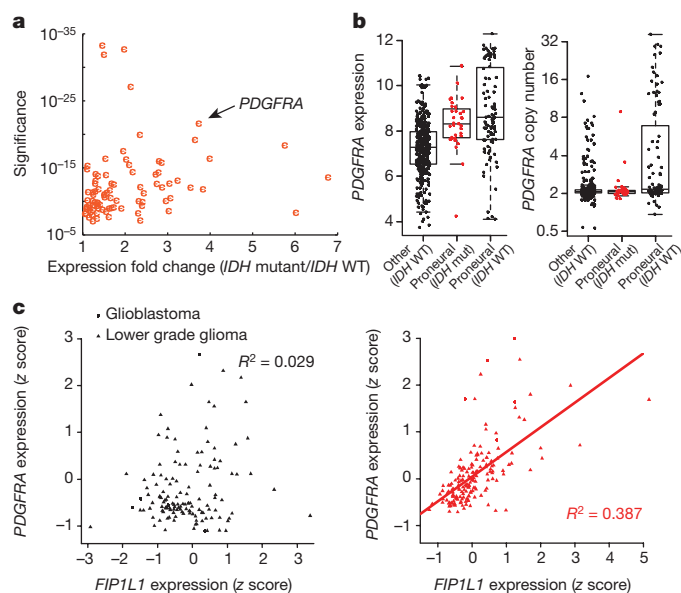




**Figure 1 | CTCF binding and gene insulation compromised in IDH mutant gliomas.** **a**, Binding profiles for the methylation-sensitive insulator CTCF are shown for a representative locus in IDH mutant and wild-type tumours, normalized by average signal. **b**, Scatterplot compares CTCF binding signals between IDH mutant (y axis) and wild-type (x axis) gliomas for all detected CTCF sites. A larger fraction of sites is commonly lost in all IDH mutants ( $n = 625$ ) than gained ( $n = 300$ ). **c**, Histogram compares GC content between CTCF sites that are lost or retained. **d**, Box plots show DNA methylation levels over lost CTCF sites, as determined by whole-genome bisulfite data for three IDH wild-type and three IDH mutant tumours. **e**, Plot depicts average Pearson correlation between gene pairs as a function of distance across RNA-seq profiles for human brain<sup>20</sup>. Gene pairs separated by a constitutive CTCF-bound boundary per HiC<sup>15</sup> have lower correlations. **f**, Volcano plot depicts the significance (y axis) of gene pairs that are more (or less) correlated in IDH mutant than in wild-type (WT) lower-grade gliomas. Gene pairs with significantly increased correlations in IDH mutants (right) tend to cross boundaries (orange), while those with decreased correlations (left) are more likely reside in the same domain (blue). These data indicate that IDH mutant, G-CIMP gliomas have reduced CTCF binding and altered expression patterns suggestive of defective gene insulation.

amplification and gain-of-function mutations in glioblastoma (15%), such alterations are rare in IDH mutant tumours<sup>23,24</sup>. Nonetheless, IDH mutant gliomas strongly express *PDGFRA* (Fig. 2b), and share the proneural transcriptional program characteristic of *PDGFRA*-amplified tumours<sup>23,24</sup>. Closer examination of the expression patterns in IDH mutant gliomas reveals a marked correlation between *PDGFRA* and *FIP1L1*, despite an intervening boundary (Fig. 2c). *FIP1L1* encodes an RNA-processing protein that is constitutively expressed in neural tissues, and particularly active in oligodendrocyte precursors, a putative glioma cell of origin<sup>22</sup> (Extended Data Fig. 2a). Moreover, combined expression of *PDGFRA* and *FIP1L1* is associated with poorer outcome in IDH mutant lower-grade gliomas (Extended Data Fig. 2b). This suggests that an aberrant interaction with this constitutive locus may drive *PDGFRA* expression in IDH mutant tumours.

We therefore investigated the topology of the region using kilobase-resolution HiC data<sup>15</sup>. In all six cell types examined, *PDGFRA* and *FIP1L1* reside in distinct domains, separated by one CTCF-anchored constitutive boundary (Fig. 3a and Extended Data Fig. 3). Our ChIP-seq data confirm that this boundary contains a strong CTCF-binding site over a canonical CTCF motif with a CpG dinucleotide

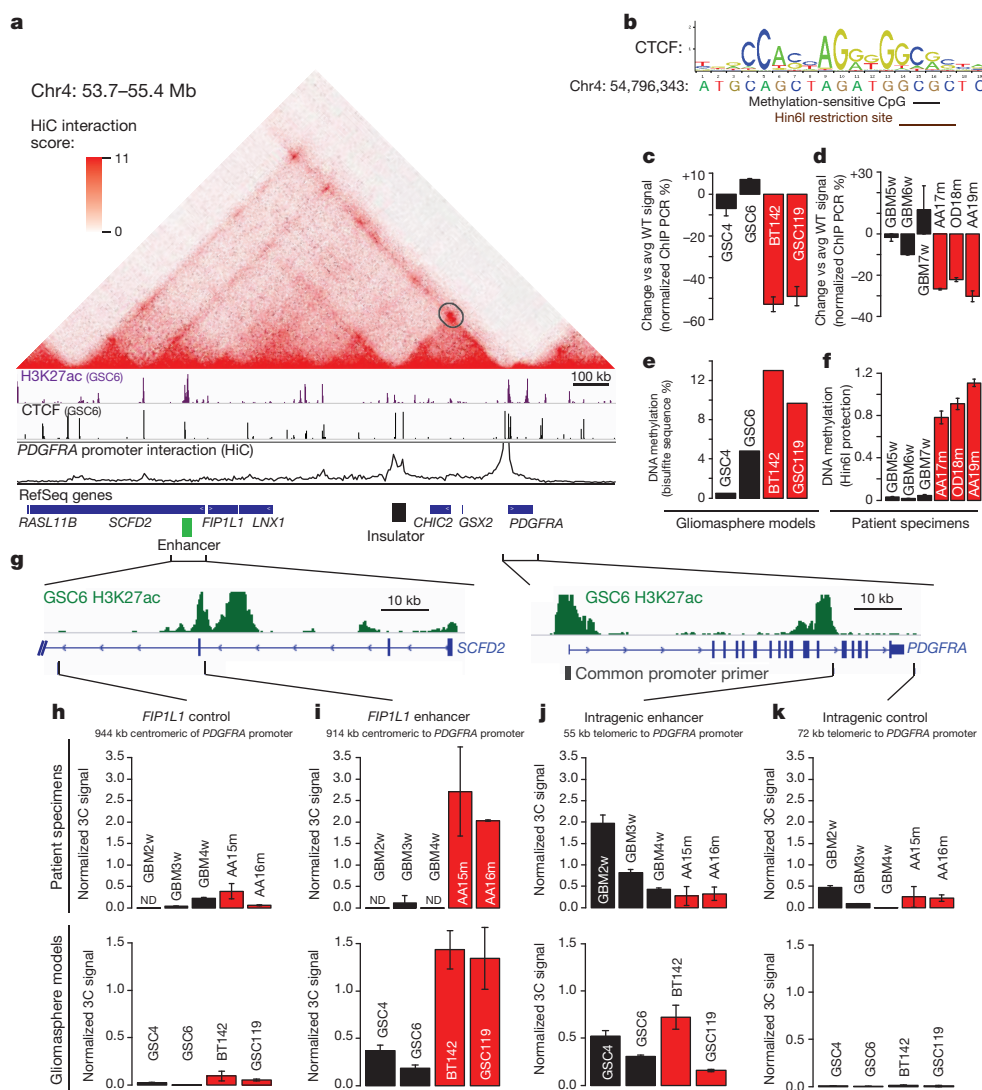


**Figure 2 | Topological domain boundaries disrupted in IDH mutant gliomas.** **a**, Scatterplot depicts significance of deregulated boundaries in IDH mutant tumours (y axis) against fold change of most upregulated gene in adjacent domains (x axis). *PDGFRA* is adjacent to a significantly deregulated boundary and upregulated in IDH mutants. **b**, Boxplots compare *PDGFRA* expression (left) or copy number (right) for 443 glioblastoma tumours, classified by IDH status and expression subtype<sup>24</sup>. IDH mutants (red) have increased *PDGFRA* expression, despite normal copy number. **c**, Plots compare *PDGFRA* (y axis) and *FIP1L1* (x axis) expression in IDH wild-type (left) and mutant (right) gliomas. The genes correlate specifically in IDH mutants, consistent with deregulation of the intervening boundary/insulator.

in a position previously linked to methylation-sensitivity<sup>25</sup> (Fig. 3b). Quantitative ChIP-PCR reveals that CTCF occupancy at this site is reduced between 30% and 50% in IDH mutant tumours and glioma-sphere models, relative to wild type (Fig. 3c, d). Moreover, the CpG in this motif becomes highly methylated in IDH mutants (Fig. 3e, f). This suggests that reduced CTCF binding may compromise the boundary flanking *PDGFRA* in IDH mutant, hypermethylated tumours.

To identify regulatory elements that might underlie *PDGFRA* induction, we mapped the enhancer-associated histone modification, histone H3 lysine 27 acetylation (H3K27ac), in glioma specimens and models. We identified a large enhancer ~50 kb upstream of *FIP1L1* with strong acetylation in wild-type and mutant tumours (Fig. 3a and Extended Data Fig. 4). In support of an enhancer identity, the element is enriched for H3 lysine 4 mono-methylation (H3K4me1), but lacks H3K4me3, and contains conserved motifs bound by the glioma master transcription factors OLIG2 and SOX2. Although this enhancer is normally insulated from *PDGFRA*, we reasoned that disruption of the intervening boundary might allow it to interact with the oncogene in IDH mutant gliomas. To test this, we used chromosome conformation capture (3C) to query the relative frequencies with which the *PDGFRA* promoter interacts with the *FIP1L1* enhancer, with an intragenic *PDGFRA* enhancer, or with nearby control sites (Fig. 3g). We fixed IDH mutant and wild-type glioma specimens and gliomaspheres, digested their chromatin with *HindIII*, and performed proximity ligation to re-ligate physically interacting DNA sequences. We used quantitative PCR (qPCR) to measure ligation frequencies between elements, normalizing against control ligations performed with bacterial artificial chromosome DNA.

In wild-type gliomas, 3C revealed a strong interaction between the *PDGFRA* promoter and its intragenic enhancer, which are ~50 kb apart (Fig. 3j, k). In contrast, the *PDGFRA* promoter does not interact with the *FIP1L1* enhancer in wild-type tumours, consistent with retention of the intervening boundary (Fig. 3h, i). However, the interaction patterns



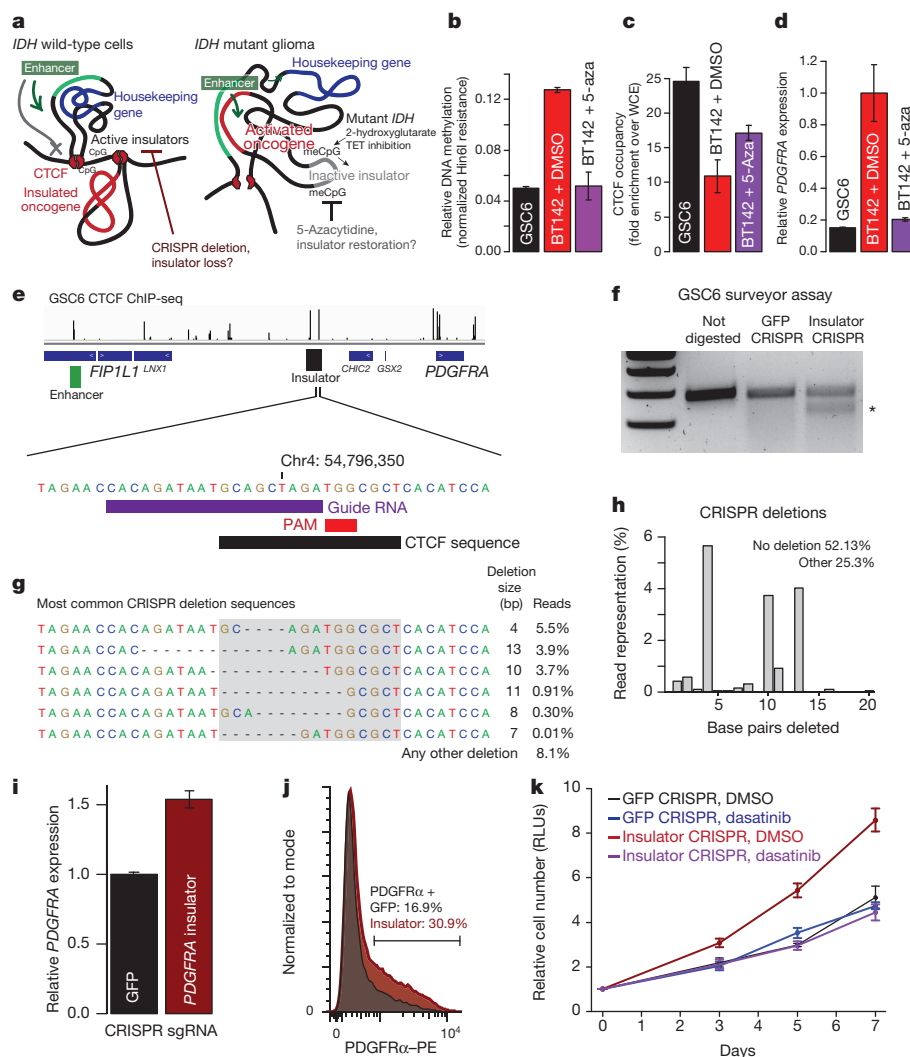
**Figure 3 | Insulator loss allows *PDGFRA* to interact with a constitutive enhancer.** **a**, Contact domain structure shown for a 1.7-Mb region containing *PDGFRA*. Heat depicts HiC interaction scores between triangulated loci in IMR90 cells<sup>15</sup>. Domains are visible as triangle-shaped regions of high interaction scores. Convergent CTCF sites anchor a loop that separates *PDGFRA* and *FIP1L1* (black circle). H3K27ac and CTCF profiles are aligned to the contact map. Interaction trace (below) depicts HiC signals between the *PDGFRA* promoter and all other positions in the region. Genes, *FIP1L1* enhancer (per H3K27ac) and insulator (per HiC and CTCF binding) are indicated. **b**, The right CTCF peak in the insulator contains a CTCF motif with a CpG at a methylation-sensitive position. **c**, **d**, ChIP-qPCR data show that CTCF occupancy over the boundary is reduced in *IDH* mutant (red) gliomas and models, relative to wild type

were markedly different in *IDH* mutant tumours. Here, 3C revealed a strong interaction between the *PDGFRA* promoter and the *FIP1L1* enhancer, despite a separation of ~900 kb (Fig. 3i). For comparison, this interaction is approximately fivefold stronger than that between the *PDGFRA* promoter and its intragenic enhancer. To confirm this interaction, we designed and normalized reciprocal probe and primers to compare the relative strength with which the *FIP1L1* enhancer interacts with nearby promoters and *PDGFRA* (Extended Data Fig. 5). Notably, we found that the interaction between *FIP1L1* enhancer and *PDGFRA* promoter in *IDH* mutant tumours is stronger than that between *FIP1L1* enhancer and *FIP1L1* promoter. This suggests that disruption of a boundary element by *IDH* mutation and hypermethylation allows a potent constitutive enhancer to interact aberrantly with, and upregulate, *PDGFRA*.

(black). **e**, Methylation levels of the CpG in the CTCF motif were measured in gliomaspheres by bisulfite sequencing, and plotted as a percentage of alleles protected from conversion. **f**, Methylation levels of the CpG in the CTCF motif were measured in glioma specimens by methylation-sensitive restriction, and plotted as relative protection. **g**, Expanded views of the *FIP1L1* enhancer locus and *PDGFRA* locus shown with H3K27ac tracks. Vertical black bars indicate the locations of the common *PDGFRA* promoter primer and four complementary primers tested in 3C. **h–k**, Plots show normalized 3C interaction frequencies between *PDGFRA* promoter and the *FIP1L1* enhancer and indicated regions. A strong interaction between the *PDGFRA* promoter and the *FIP1L1* enhancer is evident in *IDH* mutant tumours and models. ND, none detected. Bars and error bars in all panels reflect mean and s.d. of triplicate observations, respectively.

To test this model functionally, we considered whether perturbing the boundary alters *PDGFRA* expression in patient-derived gliomaspheres (Fig. 4a). First, we focused on the *IDH1* mutant astrocytoma model, BT142. In this mutant line, the CpG dinucleotide in the CTCF motif exhibits higher methylation than wild-type models (~13% versus ~2% per bisulfite sequencing), and CTCF binding is roughly threefold lower. Consistently, 3C reveals a strong interaction between the *FIP1L1* enhancer and the *PDGFRA* promoter that is specific to the mutant line (Fig. 3i), and *PDGFRA* is highly expressed.

We reasoned that a demethylating agent should reduce methylation at this CpG dinucleotide, allowing CTCF to bind and restore *PDGFRA* insulation. We therefore treated BT142 gliomaspheres with the DNA methyltransferase inhibitor 5-azacytidine (5-aza). 5-aza treatment reduced methylation of the CTCF motif by ~2.5-fold, increased CTCF



**Figure 4 | Boundary methylation and CTCF occupancy affect *PDGFRA* expression and proliferation.** **a**, Schematic depicts chromatin loops and boundaries in the *PDGFRA* locus. In *IDH* wild-type cells (left), intact boundary insulates oncogene. Disruption of the boundary by removing the CTCF motif should activate the oncogene. In *IDH* mutant cells (right), hypermethylation blocks CTCF, compromising the boundary and allowing enhancer to activate the oncogene. Demethylation should restore CTCF-mediated insulation. meCpG, methylated CpG. **b**, Plot compares CpG methylation in the CTCF motif in *IDH* wild-type gliomaspheres (black), *IDH1* mutant gliomaspheres (red), and *IDH1* mutant gliomaspheres treated with 5  $\mu$ M 5-aza for 8 days (purple). **c**, Plot compares CTCF occupancy over the boundary. DMSO, dimethylsulfoxide; WCE, whole-cell extract. **d**, Plot compares *PDGFRA* expression. Demethylation restores *PDGFRA* insulation in *IDH1* mutant gliomaspheres. **e**, CTCF binding shown for the *FIP1L1*/*PDGFRA* region. Expanded view shows CTCF motif in the

insulator targeted for CRISPR-based deletion. sgRNA and PAM direct Cas9 nuclease to the motif. **f**, Surveyor assay detects target site alterations in GSC6 gliomaspheres infected with Cas9 and sgRNA (but not in control cells infected with GFP-targeting sgRNA). **g**, Sequencing of target site reveals the indicated deletions. CTCF motif disrupted on ~25% of alleles (compare to <0.01% in control). **h**, Plot depicts fraction of reads in insulator CRISPR cells with a deletion of indicated size. **i**, qPCR reveals increased *PDGFRA* expression in insulator CRISPR cells. **j**, Flow cytometry reveals ~2-fold greater *PDGFR $\alpha$*  in insulator CRISPR cells. PE, phycoerythrin. **k**, Plot depicts gliosphere growth. Insulator CRISPR cells exhibit an approximately twofold increased proliferation, relative to control. This proliferation advantage is eliminated by *PDGFR $\alpha$*  inhibition. RLUs, relative light units. These results indicate that genetic or epigenetic disruption of the boundary compromises insulation of this oncogene. Bars and error bars in all panels reflect mean and s.d. of triplicate observations, respectively.

occupancy by ~1.7-fold and downregulated *PDGFRA* expression by ~5-fold (Fig. 4b–d). These results directly implicate DNA hypermethylation in compromising CTCF binding, boundary function and oncogene insulation in *IDH* mutant tumours.

Finally, we investigated whether genetic disruption of the CTCF motif could induce *PDGFRA* expression in wild-type gliomaspheres with an intact boundary (Fig. 4a). Here we focused on GSC6, a patient-derived glioblastoma model that contains an *EGFR* amplification, but is wild type for *IDH1*, *IDH2* and *PDGFRA*. We sought to disrupt the CTCF site in the boundary by CRISPR (clustered regularly interspaced short palindromic repeats)-based genome engineering<sup>26,27</sup> (Fig. 4e). We designed a short guide RNA (sgRNA) with a protospacer adjacent motif (PAM) within the CTCF motif. A single-vector lentiviral delivery system was used to infect GSC6 cells with a Cas9 expression

construct containing this insulator sgRNA or a control sgRNA (targeting green fluorescent protein, GFP). Surveyor assay confirmed target locus disruption in the insulator CRISPR condition (Fig. 4f). Direct sequencing of the target locus revealed that ~25% of alleles in the insulator CRISPR gliomaspheres contain a deletion within the CTCF motif expected to disrupt binding, compared to <0.1% in the GFP control (Fig. 4g, h).

We quantified *PDGFRA* expression in the genetically modified gliomaspheres. Reverse transcription PCR (RT-PCR) revealed an ~1.6-fold increase in *PDGFRA* messenger RNA in the insulator CRISPR cells, relative to control (Fig. 4i). Similarly, flow cytometry revealed an ~1.8-fold increase in the fraction of cells with *PDGFR $\alpha$*  surface expression (Fig. 4j). We conservatively estimate that CTCF motif disruption causes an ~3-fold increase in *PDGFRA* expression, given that



DNA level analysis indicates that less than 50% of insulator CRISPR cells were successfully edited.

Finally, we considered whether CRISPR-mediated boundary disruption and *PDGFRA* induction affects gliomasphere fitness. In support, the insulator CRISPR gliomaspheres have an approximately two-fold growth advantage over the control GFP CRISPR gliomaspheres (Fig. 4k). This growth advantage is dependent on PDGFR $\alpha$  signalling, as it is abrogated by treatment with the PDGFR inhibitors dasatinib or crenolanib (Fig. 4k and Extended Data Fig. 6). Notably, *PDGFRA* expression in insulator CRISPR gliomaspheres increased further after extended culture (twofold increase compared with control), potentially owing to selection of effectively edited clones. The observation that genetic disruption of this CTCF boundary element induces *PDGFRA* expression and enhances proliferation provides strong support for our model that epigenetic disruption of this element offers similar growth advantage to *IDH* mutant gliomas.

In conclusion, we present a new epigenetic mechanism by which gain-of-function *IDH* mutations induce *PDGFRA* expression and thereby promote fitness in a subset of gliomas. We specifically find that, in addition to familiar effects on CpG islands, *IDH* mutations cause hypermethylation of CTCF binding sites genome-wide. This is associated with reduced CTCF binding and a global deregulation of boundary elements that partition topological domains. Disruption of a specific boundary bordering *PDGFRA* allows a potent enhancer to contact and activate this canonical glioma oncogene aberrantly.

Although disruption of this single boundary confers a growth advantage, it is unlikely to be the only mediator of *IDH* mutations in gliomas. The widespread disruption of CTCF binding and boundary element function could provide many opportunities for oncogene deregulation, and subsequent selection of proliferative progeny that inherit the altered epigenetic state. Insulator dysfunction may also be accompanied by promoter silencing events<sup>28,29</sup>, and by alterations to other pathways affected by 2-hydroxyglutarate<sup>7,30</sup>. Conversely, disruption of chromosomal topology and oncogene insulation may be more generally relevant to methylator phenotypes observed in colorectal and renal cell carcinomas, leukaemia and other malignancies<sup>28</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 July; accepted 26 November 2015.

Published online 23 December 2015.

1. Parsons, D. W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807–1812 (2008).
2. The Cancer Genome Atlas Research Network. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N. Engl. J. Med.* **372**, 2481–2498 (2015).
3. Dang, L. *et al.* Cancer-associated *IDH1* mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–744 (2009).
4. Figueroa, M. E. *et al.* Leukemic *IDH1* and *IDH2* mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**, 553–567 (2010).
5. Xu, W. *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of  $\alpha$ -ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17–30 (2011).
6. Lu, C. *et al.* *IDH* mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* **483**, 474–478 (2012).
7. Cairns, R. A. & Mak, T. W. Oncogenic isocitrate dehydrogenase mutations: mechanisms, models, and clinical opportunities. *Cancer Dis.* **3**, 730–741 (2013).
8. Pastor, W. A., Aravind, L. & Rao, A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nature Rev. Mol. Cell Biol.* **14**, 341–356 (2013).

9. Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502**, 472–479 (2013).
10. Noshahr, H. *et al.* Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17**, 510–522 (2010).
11. Turcan, S. *et al.* *IDH1* mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* **483**, 479–483 (2012).
12. Bickmore, W. A. & van Steensel, B. Genome architecture: domain organization of interphase chromosomes. *Cell* **152**, 1270–1284 (2013).
13. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
14. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
15. Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
16. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
17. Lupiáñez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
18. Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* **405**, 482–485 (2000).
19. Hark, A. T. *et al.* CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. *Nature* **405**, 486–489 (2000).
20. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
21. Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl Acad. Sci. USA* **111**, 996–1001 (2014).
22. Sturm, D. *et al.* Paediatric and adult glioblastoma: multifactorial (epi)genomic culprits emerge. *Nature Rev. Cancer* **14**, 92–107 (2014).
23. Brennan, C. W. *et al.* The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477 (2013).
24. Verhaak, R. G. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in *PDGFRA*, *IDH1*, *EGFR*, and *NF1*. *Cancer Cell* **17**, 98–110 (2010).
25. Wang, H. *et al.* Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* **22**, 1680–1688 (2012).
26. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).
27. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnol.* **32**, 347–355 (2014).
28. Baylín, S. B. & Jones, P. A. A decade of exploring the cancer epigenome – biological and translational implications. *Nature Rev. Cancer* **11**, 726–734 (2011).
29. Costello, J. F., Berger, M. S., Huang, H. S. & Cavennee, W. K. Silencing of *p16/CDKN2* expression in human gliomas by methylation and chromatin condensation. *Cancer Res.* **56**, 2405–2410 (1996).
30. Koivunen, P. *et al.* Transformation by the (R)-enantiomer of 2-hydroxyglutarate linked to EGLN activation. *Nature* **483**, 484–488 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank J. Kim, the MGH Neuro Oncology Tissue Repository, and the MGH Pathology Flow Cytometry Core for assistance with clinical samples and analysis, and E. Lander and W. Kaelin for discussions. W.A.F. is supported by a basic research fellowship from the American Brain Tumor Association. B.B.L. is supported by a Jane Coffin Childs fellowship. B.E.B. is an American Cancer Society Research Professor. This research was supported by funds from Howard Hughes Medical Institute, the National Brain Tumor Society and the National Human Genome Research Institute.

**Author Contributions** Conception and experimental design: W.A.F., Y.D., B.B.L., S.M.G., M.L.S. and B.E.B. Methodology and data acquisition: W.A.F., Y.D., B.B.L., S.M.G., A.S.V., A.O.S.-R., M.L.S. and B.E.B. Analysis and interpretation of data: W.A.F., Y.D. and B.E.B. Manuscript writing: W.A.F., Y.D. and B.E.B. W.A.F. and Y.D. contributed equally to this work.

**Author Information** Data generated for this study are available through the Gene Expression Omnibus (GEO) under accession number GSE70991. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.E.B. ([bernstein.bradley@mgh.harvard.edu](mailto:bernstein.bradley@mgh.harvard.edu)).



## METHODS

No statistical methods were used to predetermine sample size.

**Primary glioma specimens and gliomasphere models.** Clinical samples GBM1w, GBM2w, GBM3w, GBM4w, GBM5w, GBM6w, GBM7w, AA15m, AA16m, AA17m, OD18m and AA19m were obtained as frozen specimens from the Massachusetts General Hospital Pathology Tissue Bank, or received directly after surgical resection and flash frozen (Extended Data Table 1). All samples were acquired with Institutional Review Board approval, and were de-identified before receipt. GBM1w was obtained at autopsy; the remaining samples were surgical resections. *IDH* status was determined for all clinical samples by SNaPshot multiplex PCR<sup>31</sup>. *PDGFRA* status was confirmed by FISH analysis. Tissue (200–500 µg) was mechanically minced with a sterile razor blade before further processing.

Gliomaspheres were maintained in culture as described<sup>32,33</sup>. In brief, neurosphere cultures contain Neurobasal media supplemented with 20 ng ml<sup>-1</sup> recombinant EGF (R and D Systems), 20 ng ml<sup>-1</sup> FGF2 (R and D Systems), 1 × B27 supplement (Invitrogen), 0.5 × N2 supplement (Invitrogen), 3 mM L-glutamine, and penicillin/streptomycin. Cultures were confirmed to be mycoplasma-free via PCR methods. GSC4 and GSC6 gliomasphere lines were derived from *IDH* wild-type tumours resected at Massachusetts General Hospital, and have been previously described and characterized<sup>32–34</sup>. BT142 gliomasphere line (*IDH1* mutant)<sup>35</sup> was obtained from ATCC, and cultured as described above except 25% conditioned media was carried over each passage. BT142 G-CIMP status was confirmed by evaluating LINE methylation with the Global DNA Methylation Assay – LINE-1 kit (Active Motif), as described<sup>36</sup>, and by methylation-sensitive restriction digests. GSC119 was derived from an *IDH1* mutant tumour (confirmed by SNaPshot) resected at Massachusetts General Hospital. We confirmed *IDH1* mutant status of GSC119 by RNA-seq (82 out of 148 reads overlapping the relevant position in the transcript correspond the mutant allele). The gliomasphere models were derived from tumours of the following types: GSC4 and GSC6: primary glioblastoma; BT142: grade III oligoastrocytoma; GSC119: secondary glioblastoma, G-CIMP. Clinical specimens and models used in this study are detailed in Extended Data Table 1.

**ChIP.** ChIP-seq was performed as described previously<sup>32</sup>. In brief, cultured cells or minced tissue was fixed in 1% formaldehyde and snap frozen in liquid nitrogen and stored at –80 °C at least overnight. Sonication of tumour specimens and gliomaspheres was calibrated such that DNA was sheared to between 400 and 2,000 bp. CTCF was immunoprecipitated with a monoclonal rabbit CTCF antibody, clone D31H2 (Cell Signaling 3418). H3K27ac was immunoprecipitated with an antibody from Active Motif (39133). ChIP DNA was used to generate sequencing libraries by end repair (End-It DNA repair kit, Epicentre), 3' A base overhang addition via Klenow fragment (NEB), and ligation of barcoded sequencing adapters. Barcoded fragments were amplified by PCR. Libraries were sequenced as 38-base paired-end reads on an Illumina NextSeq500 instrument or as 50-base single-end reads on a MiSeq instrument. Sequencing libraries are detailed in Extended Data Table 2. H3K27ac maps for GSC6 were previously deposited to the GEO under accession GSM1306340. Genomic data has been deposited into GEO as GSE70991.

For sequence analysis, identical reads were collapsed to a single paired-end read to avoid PCR duplicates. To avoid possible saturation, reads were downsampled to 5% reads collapsed as PCR duplicates, or 5 million fragments. Reads were aligned to hg19 using BWA, and peaks were called using HOMER. ChIP-seq tracks were visualized using Integrative Genomics Viewer (IGV, <http://www.broadinstitute.org/igv/>). To detect peaks lost in *IDH* mutants, we called signal over all peaks in a 100-bp window centred on the peaks. To control for copy number changes, we first called copy number profiles from input sequencing data using CNVator<sup>37</sup>. We then removed all regions where at least one sample had a strong deletion (<0.25), and normalized by copy number. To account for batch effects and difference in ChIP efficiency, we quantile normalized each data set. Peaks were scored as lost or gained if the difference in signal between a given tumour and the average of the five wild-type tumours was at least twofold lower or higher, with a signal of at least 1 in all wild-type or *IDH* mutant tumours. Fisher exact test confirmed that the overlap between peaks lost in the *IDH* mutant tumours is highly significant ( $P < 10^{-100}$ ).

GC content over CTCF peaks lost (or retained) in the *IDH* mutant glioma specimens was averaged over 200-bp windows centred on each peak lost in *IDH* mutant tumours. Methylation levels were quantified over these same regions for 3 *IDH* mutant and 3 *IDH* wild-type tumours, using TCGA data generated by whole genome bisulfite sequencing<sup>10</sup>. In brief, methylation levels (percentage) based on proportion of reads with protected CpG were averaged over all CpG di-nucleotides in these regions, treating each tumour separately.

Occupancy of the CTCF site in the boundary element adjacent to the *PDGFRA* locus was quantified by ChIP qPCR, using the following primers: *PDGFRA*ctcf: 5'-GTCACAGTAGAACCACAGAT-3'; *PDGFRA*ctcfR: 5'-TAAGTATACTGGTCCTCCTC-3'. Equal masses of ChIP or input (WCE) DNA were used as input for PCR, and CTCF occupancy was quantified as a ratio between ChIP and WCE, determined by  $2^{-\Delta\Delta C_t}$ . CTCF peak intensity was further

normalized as ratio to two invariant peaks, at *PSMB1* and *SPG11*, using the following primers: *PSMB1*ctcf: 5'-CCTTCCTAGTCACTCAGTAA-3'; *PSMB1*ctcfR: 5'-CAGTGTGACTCATCCAG-3'; *SPG11*ctcf: 5'-CAGTACCAGCCTCTCTAG-3'; *SPG11*ctcfR: 5'-CTAAGCTAGGCCTTCAAG-3'.

**Cross-boundary and intra-domain gene pair correlation analysis.** RNA-seq data for 357 normal brain samples was downloaded from GTEx<sup>20</sup>. RNA-seq data and copy number profiles for lower grade gliomas were downloaded from TCGA<sup>23,24</sup>. Contact domains of IMR90, GM12878, K562 and NHEK cells were obtained from published HiC data<sup>15</sup>. Genes were assigned to the inner-most domain in which their transcription start site fell within. Gene pairs were considered to be in the same domain if they were assigned to the same domain in both GM12878 and IMR90. Gene pairs were considered to span a boundary if they were assigned to different domains in both GM12878 and IMR90, and separated by a CTCF-binding site in *IDH* wild-type tumours. Gene pairs that did not fit either criterion were excluded from this analysis. The plot of correlation vs distance for brain GTEx samples is based on Pearson correlations for all relevant pairs, smoothed by locally weighted scatterplot smoothing with weighted linear least squares (LOESS). To assess the bias in correlation differences, we computed the difference of Pearson correlations between wild-type and *IDH* mutant gliomas for all gene pairs separated by <180 kb. In Fig. 1e, this difference in correlations is plotted against the significance of this difference (estimated by Fisher  $z$ -transformation). For each gene pair, we omitted samples with a deletion or amplification of one of the genes at or above threshold of the minimal arm level deletion or amplification (to avoid copy number bias). To ensure robustness, we also repeated the analysis using boundaries defined from HiC data for K562 and NHEK. This yielded similar results: 84% pairs gaining correlation cross boundary versus 71% expected ( $P < 8 \times 10^{-3}$ ), 54% pairs losing correlation are within the same domain versus 29% expected ( $P < 3 \times 10^{-8}$ ). Repeating the analysis with only the 14,055 genes that have expressed over 1 transcripts per million (TPM) in at least half the samples also yielded similar results (Extended Data Fig. 7): 92% pairs gaining correlation cross boundary versus 69% expected ( $P < 2 \times 10^{-3}$ ), 73% pairs losing correlation are within the same domain versus 31% expected ( $P < 8 \times 10^{-4}$ ).

**Genomic scan for deregulated boundaries.** To detect boundaries deregulated in *IDH* mutant gliomas, we scanned for gene pairs, separated by <1 Mb, with a significant difference in correlation between wild-type and *IDH* mutant tumours (Fisher  $z$ -transformation, FDR <1%). We omitted amplified or deleted samples as described above. To ensure robustness to noise from lowly expressed genes, we first filtered out 6,476 genes expressed <1 TPM in more than half of the samples (keeping 14,055 genes). We considered all domains and boundaries scored in IMR90 HiC data<sup>13</sup>. Gene pairs crossing a CTCF peak and an IMR90 boundary (that is, can be assigned to different domains) that were significantly more correlated in *IDH* mutant tumours were considered to support the loss of that boundary. Gene pairs not crossing a boundary (that is, can be assigned to the same domain) that were significantly less correlated in *IDH* mutant tumours were considered to support the loss of a flanking boundary. We collated a set of deregulated boundaries, supported by at least one cross-boundary pair gaining correlation and at least one intra-domain pair losing correlation. Each was assigned a  $P$  value equal to the product of both supporting pairs (best  $P$  value was chosen if there were more supporting pairs). If both boundaries of a domain were deregulated, or if the same pair of gene pairs (one losing and one gaining correlations) were supporting more than one boundary due to overlapping domains, the entries were merged (Supplementary Table 1). This definition allows every gene pair to be considered as potential support for a boundary loss. To quantify CTCF occupancy over these deregulated boundaries, we averaged the signal over all CTCF peaks located within a 1-kb window around the boundary, using copy number and quantile normalized CTCF signals. To quantify DNA methylation over the deregulated boundaries, we averaged DNA methylation signals from TCGA data in 200-bp windows as above. Figure 2a depicts significance of disrupted domains and the fold change of genes in them that are upregulated in *IDH* mutant tumours (compared to median expression in wild type). In addition to *PDGFRA*, top-ranking genes include *CHD4* ( $P < 10^{-32}$ ), a driver of glioblastoma tumour initiation<sup>38</sup>, *LICAM* ( $P < 10^{-8}$ ), a regulator of the glioma stem cells and tumour growth<sup>39</sup>, and other candidate regulators (Supplementary Table 1).

To ensure robustness to cell-type-specific boundaries, we repeated the analysis with GM12878-, K562- and NHEK-defined boundaries. This yielded very similar results, and again highlighted *PDGFRA* as an overexpressed gene adjacent to a disrupted boundary.

**TCGA correlation and outcome analysis.** For the correlation of *FIP1L1* and *PDGFRA* expression, RNA-seq data from the TCGA lower grade glioma (LGG) and glioblastoma (GBM) data sets<sup>2,24</sup> were downloaded and segregated by *IDH* mutation status and subtype. Patients from the proneural subtype were divided by *IDH* mutation status, while patients from the mesenchymal, classical or neural subtypes (which had no *IDH* mutations) were classified as 'other'. For correlation

analysis, patients with copy number variation in either gene were excluded from the analysis to control for effects of co-amplification. For outcome analysis, LGG RNA-seq data and corresponding patient survival data was obtained from TCGA. Patients with sum *PDGFRA* and *FIP1L1* expression of at least one-half of one standard deviation above the mean were classified as 'high *PDGFRA* and *FIP1L1* expression' ( $n = 17$ ), while all other patients were classified as 'low *PDGFRA* and *FIP1L1* expression' ( $n = 201$ ). Data were plotted as Kaplan–Meier curves and statistically analysed via log–rank test.

**HiC data analysis and visualization.** HiC data<sup>15</sup> were downloaded from GEO. 5-kb resolution intra-chromosomal contact scores for chromosome 4 for the cell lines IMR90, NHEK, KBM7, K562, HUVEC, HMEC and GM12878 were filtered to the region between 53,700 and 55,400 kb. The average interaction score at each coordinate pair for all cell lines was calculated and used to determine putative insulator elements as local maxima at the interaction point of two domain boundaries. To determine the interactions of the *PDGFRA* promoter, the interaction scores of all points in the region with the *PDGFRA* promoter (chr4: 55,090,000) were plotted as a one-dimensional trace. To view the topological domain structure of the region, HiC interaction scores were visualized using Juicebox (<http://www.aidenlab.org/juicebox/>)<sup>15</sup>. Data shown is from the IMR90 cell line at 5-kb resolution, normalized to coverage.

**DNA methylation quantification.** DNA methylation was analysed in two ways. For gliomaspheres, genomic DNA was isolated via QiaAmp DNA minikit (Qiagen) and subjected to bisulfite conversion (EZ DNA Methylation Gold Kit, Zymo Research). Bisulfite-converted DNA specific to the CTCF-binding site (defined by JASPAR<sup>40</sup>) in the boundary adjacent to *PDGFRA* was amplified using the following primers forward: 5'-GAATTATAGATAATGTAGTTAGATGG-3', reverse: 5'-AAATATACTAATCCTCTCTCCAAA-3'. Amplified DNA was used to prepare a sequencing library, which was sequenced as 38-base paired-end reads on a NextSeq500. For tumours, limiting DNA yields required an alternative strategy for methylation analysis. Tumour genomic DNA was isolated from minced frozen sections of tumours by QiaAmp DNA minikit (Qiagen). Genomic DNA was digested using the methylation-sensitive restriction enzyme Hin6I (Thermo) recognizing the restriction site GCGC, or subjected to mock digestion. Protected DNA was quantified by PCR using the following primer set: *PDGFRA*insF: 5'-CGTGAGCTGAATTGTGCCTG-3', *PDGFRA*insR: 5'-TGGGAGGACAGTTTAGGGCT-3', normalizing to mock digestion.

**3C analysis.** 3C analysis was performed using procedures as described previously<sup>41,42</sup>. In brief, ~10 million cell equivalents from minced tumour specimens or gliosphere cultures were fixed in 1% formaldehyde. Fixed samples were lysed in lysis buffer containing 0.2% PMSF using a Dounce pestle. Following lysis, samples were digested with HindIII (NEB) overnight on a thermomixer at 37°C rotating at 950 r.p.m. Diluted samples were ligated using T4 DNA ligase (NEB) at 16°C overnight, followed by RNase and proteinase K treatment. DNA was extracted via phenol/chloroform/isoamyl alcohol (Invitrogen). DNA was analysed via TaqMan PCR using ABI master mix. Primers and probe were synthesized by IDT with the following sequences: common *PDGFRA* promoter: 5'-GGTCGTGCCTTTGTTTT-3'; *FIP1L1* control: 5'-CAGGGAAGAGAGGAAGTTT-3'; *FIP1L1* enhancer: 5'-TTAAGTAAGCAGGTAACTACAT-3'; intragenic enhancer: 5'-AGCC TTTGCCTCCTTTT-3'; intragenic control: 5'-CCACAGGGAGAAGGAAAT-3'; intact promoter: 5'-CAAGGAATTCGTAGGGTTC-3'; probe: 5'-/56-FAM/TTGTATGCG/ZEN/AGATAGAAGCCAGGGCAA/3IABkFQ/-3'. For the reciprocal *FIP1L1* enhancer interaction interrogation, the following primer sequences were used: common enhancer primer (as *FIP1L1* enhancer primer above): 5'-TTAAGTAAGCAGGTAACTACAT-3', *PDGFRA* promoter (as common *PDGFRA* promoter above): 5'-GGTCGTGCCTTTGTTTT-3'; *SCFD2* promoter: 5'-AATACATGGTCATGATGCTC-3'; *FIP1L1* promoter: 5'-AGGCATTGCTTAAACATAAC-3'; *FIP1L1* control: 5'-TTATTTGTAGT AGAGTTTACTGG-3'; *PDGFRA* control: 5'-ATGATAACACCACCATTCAG-3'; *FIP1L1* enhancer probe: 5'-/56-FAM/TATCCCAAC/ZEN/CAAATACAGGG CTTGG/3IABkFQ/-3'. To normalize primer signals, bacterial artificial chromosome (BAC) clones CTD-2022B5 and RP11-626H4 were obtained from Invitrogen. BAC DNA was purified via BACMAX DNA Purification kit (Epicentre) and quantified using two primer sets specific to the Chloramphenicol resistance gene: 1F: 5'-TTCGTCTCAGCCAATCCCTG-3'; 1R: 5'-TTTGGCCATG GTGAAAACGG-3'; 2F: GGTTCATCATGCCGTTTGTG-3'; 2R: 5'-CAACTCAT CGCAGTACTGTTG-3'. BAC DNA was subjected to a similar 3C protocol, omitting steps related to cell lysis, proteinase or RNase treatment. PCR signal from tumour and gliosphere 3C was normalized to digestion efficiency and BAC primer signal.

**Treatment with demethylating agent.** BT142 cells were cultured in either 5  $\mu$ M 5-azacytidine or equivalent DMSO (1:10,000) for 8 days, with drug refreshed every 2 days.

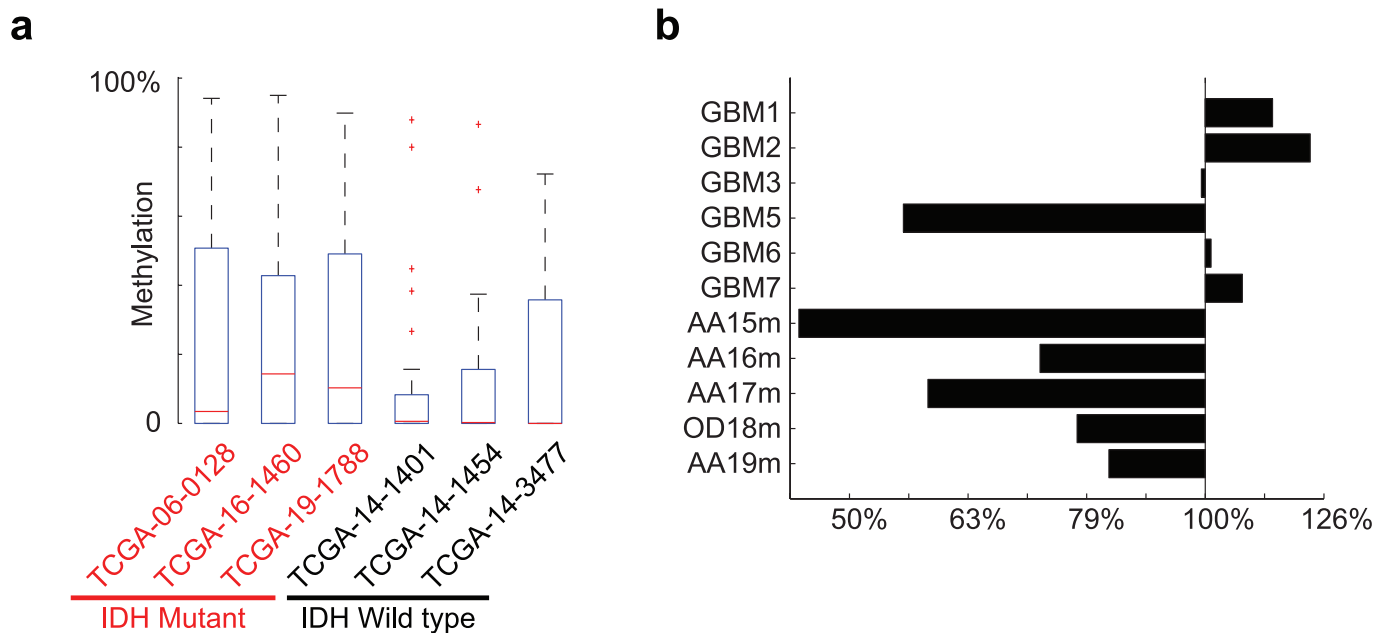
**CRISPR/Cas9 insulator disruption.** The following CRISPR sgRNAs were cloned into the LentiCRISPR vector obtained from the Zhang laboratory<sup>43</sup>: GFP: 5'-GAGCTGGACGGCGACGTAAA-3'; insulator: 5'-GCCACA GATAATGCAGCTAGA-3'. GSC6 gliomaspheres were mechanically dissociated and plated in 5  $\mu$ g ml<sup>-1</sup> EHS laminin (Sigma) and allowed to adhere overnight, and then infected with lentivirus containing either CRISPR vector for 48 h. Cells were then selected in 1  $\mu$ g ml<sup>-1</sup> puromycin for 4 days, with puromycin-containing media refreshed every 2 days. Genomic DNA was isolated and the region of interest was amplified using the *PDGFRA*ins primer set described above. CRISPR-mediated disruption of this amplified DNA was confirmed via Surveyor Assay (Transgenomic), with amplified uninfected GSC6 genomic DNA being added to each annealing reaction as the unmodified control. To quantify the precise CRISPR alterations, genomic DNA from each construct was amplified using a set of primers closer to the putative deletion site as follows: forward: 5'-TTTGCAATGGGACACGGAGA-3', reverse: 5'-AGAAATGTGTGGATGTGAGCG-3'. PCR product from these primers was used to prepare a library that was sequenced as 38-base paired-end reads on the Illumina NextSeq500.

***PDGFRA* qPCR.** Total RNA was isolated from CRISPR-infected GSC6 gliomaspheres (insulator or control GFP sgRNA) or BT142 gliomaspheres (5-aza-treated or control condition) using the RNeasy minikit (Qiagen) and used to synthesize cDNA with the SuperScriptIII system (Invitrogen). cDNA was analysed using SYBR mastermix (Applied Biosystems) on a 7500 Fast Real Time System (Applied Biosystems). *PDGFRA* expression was determined using the following primers: forward: 5'-GCTCAGCCCTGTGAGAAGAC-3', reverse: 5'-ATTGCGGAATAACATCGGAG-3', and was normalized to primers for ribosomal protein, large, P0 (*RPLP0*), as follows: forward: 5'-TCCCACCTTGCTGAAAAGGTCA-3', reverse: 5'-CCGACTCTCTCTTG GCTTCA-3'. Normalization was also verified by  $\beta$ -actin (*ACTB*), forward: 5'-AGAAAATCTGGCACCACACC-3', reverse: 5'-AGAGGCGTACAGG GATAGCA-3'.

***PDGFR $\alpha$*  flow cytometry.** Cells were incubated with PE-conjugated anti-*PDGFR $\alpha$*  (CD140a) antibody (Biolegend, clone 16A1) for 30 min at room temperature at the dilution specified in the manufacturer's protocol. Data was analysed and visualized with FlowJo software. Single live cells were selected for analysis via side and forward scatter, and viable cells were selected by lack of an unstained channel (APC) autofluorescence.

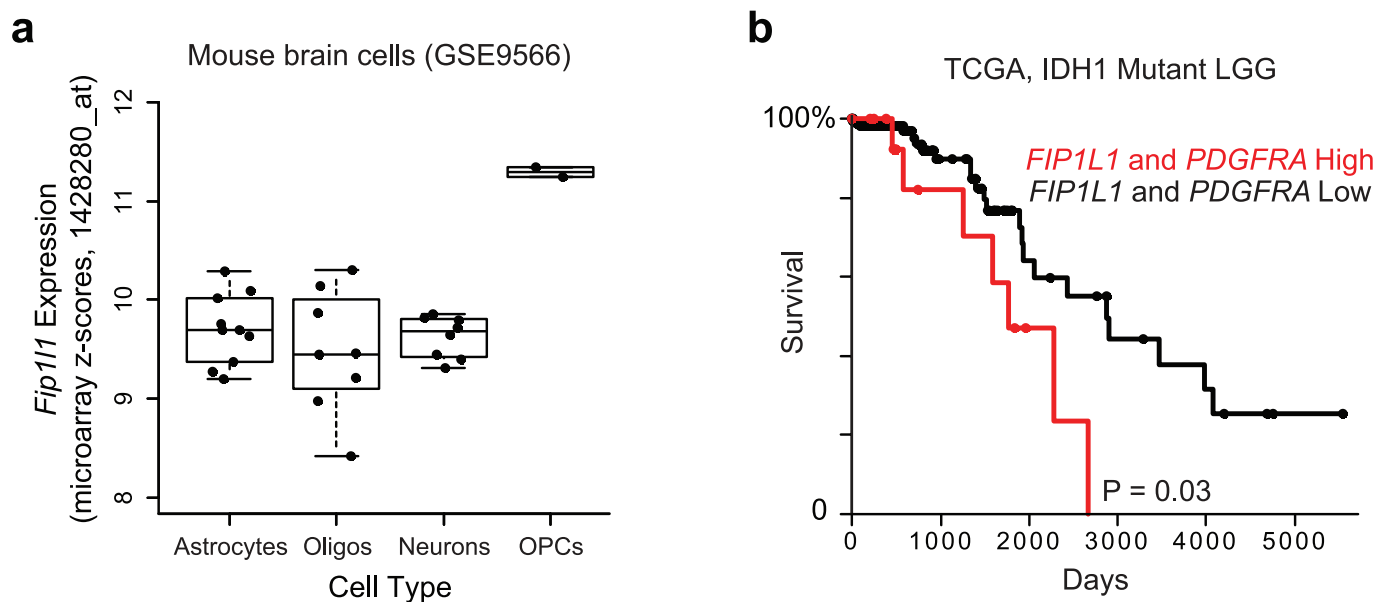
**Cell growth assay.** For the cell growth assay, 2,500 dissociated viable GSC6 cells expressing CRISPR and either GFP or insulator-targeting sgRNA (see above) was plated in 100  $\mu$ l of media in an opaque-walled tissue culture 96-well plate, in 1  $\mu$ M dasatinib, 500 nM crenolanib, or equivalent DMSO (1:10,000) as a vehicle control. Cell growth was analysed at days 3, 5 and 7 for dasatinib, or days 3, 7 and 10 for crenolanib, using CellTiter-Glo reagent (Promega) following the manufacturer's protocol. Data were normalized across days using an ATP standard curve.

- Chi, A. S. *et al.* Prospective, high-throughput molecular profiling of human gliomas. *J. Neurooncol.* **110**, 89–98 (2012).
- Rheinbay, E. *et al.* An aberrant transcription factor network essential for Wnt signaling and stem cell maintenance in glioblastoma. *Cell Rep.* **3**, 1567–1579 (2013).
- Suvà, M. L. *et al.* Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell* **157**, 580–594 (2014).
- Wakimoto, H. *et al.* Maintenance of primary tumor phenotype and genotype in glioblastoma stem cells. *Neuro Oncol.* **14**, 132–144 (2012).
- Luchman, H. A. *et al.* An *in vivo* patient-derived model of endogenous *IDH1*-mutant glioma. *Neuro Oncol.* **14**, 184–191 (2012).
- Lai, R. K. *et al.* Genome-wide methylation analyses in glioblastoma multiforme. *PLoS ONE* **9**, e89376 (2014).
- Abyzov, A., Urban, A. E., Snyder, M. & Gerstein, M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* **21**, 974–984 (2011).
- Chudnovsky, Y. *et al.* ZFH4 interacts with the NuRD core member CHD4 and regulates the glioblastoma tumor-initiating cell state. *Cell Rep.* **6**, 313–324 (2014).
- Bao, S. *et al.* Targeting cancer stem cells through L1CAM suppresses glioma growth. *Cancer Res.* **68**, 6043–6048 (2008).
- Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W. W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* **32**, D91–D94 (2004).
- de Laat, W. & Dekker, J. 3C-based technologies to study the shape of the genome. *Methods* **58**, 189–191 (2012).
- Hagège, H. *et al.* Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nature Protocols* **2**, 1722–1733 (2007).
- Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
- Cahoy, J. D. *et al.* A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28**, 264–278 (2008).



**Extended Data Figure 1 | DNA methylation and CTCF binding at deregulated boundaries.** **a**, Box plots show DNA methylation levels over CTCF sites (200-bp window centred on the peak) within boundaries predicted by gene pair correlation analysis to be disrupted. All CTCF sites located within a 1-kb window centred on a disrupted boundary were

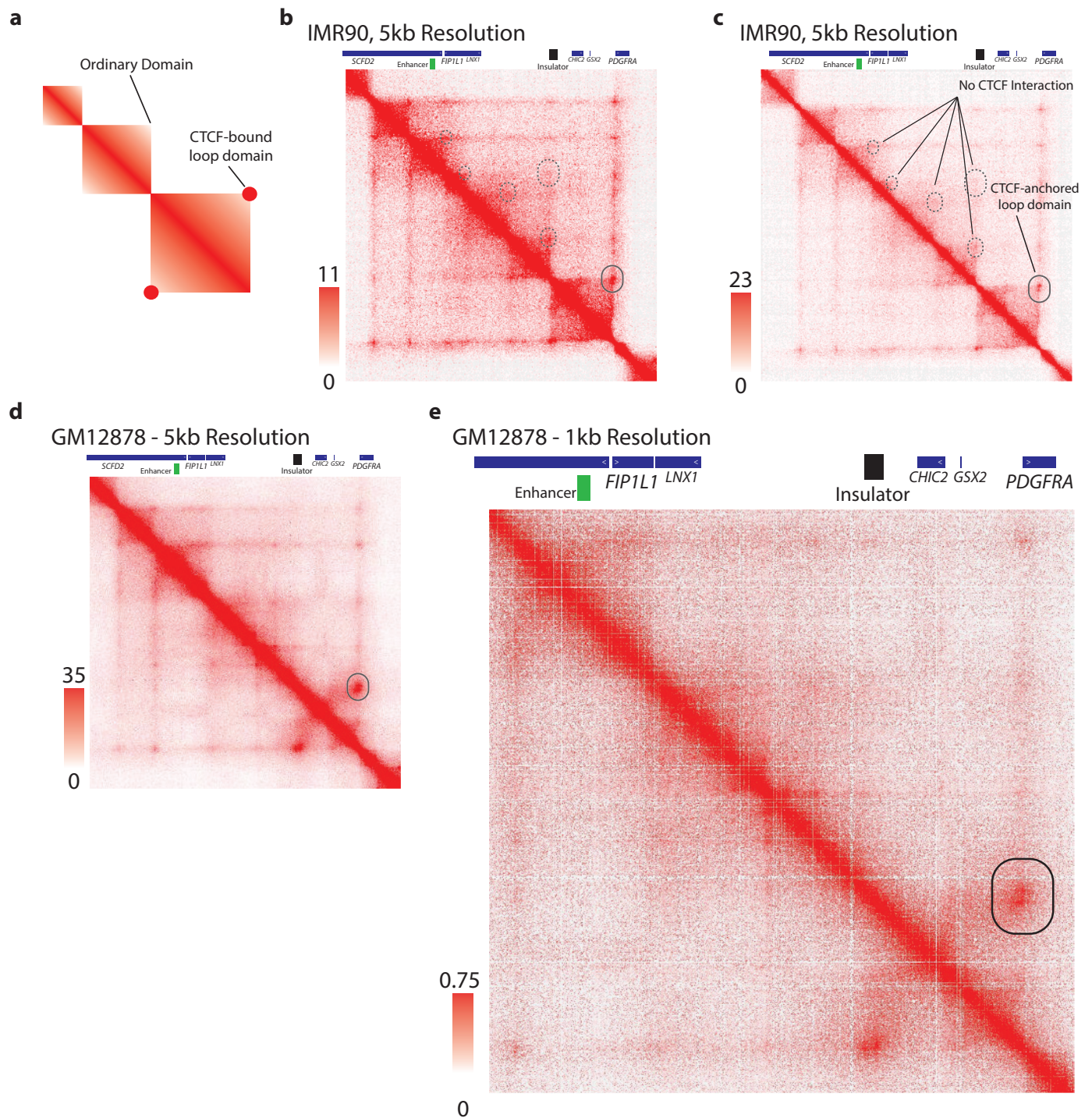
considered. Methylation levels were determined from whole-genome bisulfite data for three *IDH* mutant (red labels) and three *IDH* wild-type (black labels) tumours. **b**, Bars show average normalized ChIP-seq signal over all CTCF sites located inside a 1-kb window centred on a disrupted boundary.



**Extended Data Figure 2 | Expression of *Fip1l1* in mouse brain cells and survival effects of *PDGFRA* and *FIP1L1*.** **a**, Expression of *Fip1l1* in a published data set for isolated mouse brain cell types<sup>44</sup>. **b**, Kaplan–Meier plot based on TCGA data<sup>23</sup> indicates that combined *FIP1L1* and *PDGFRA*

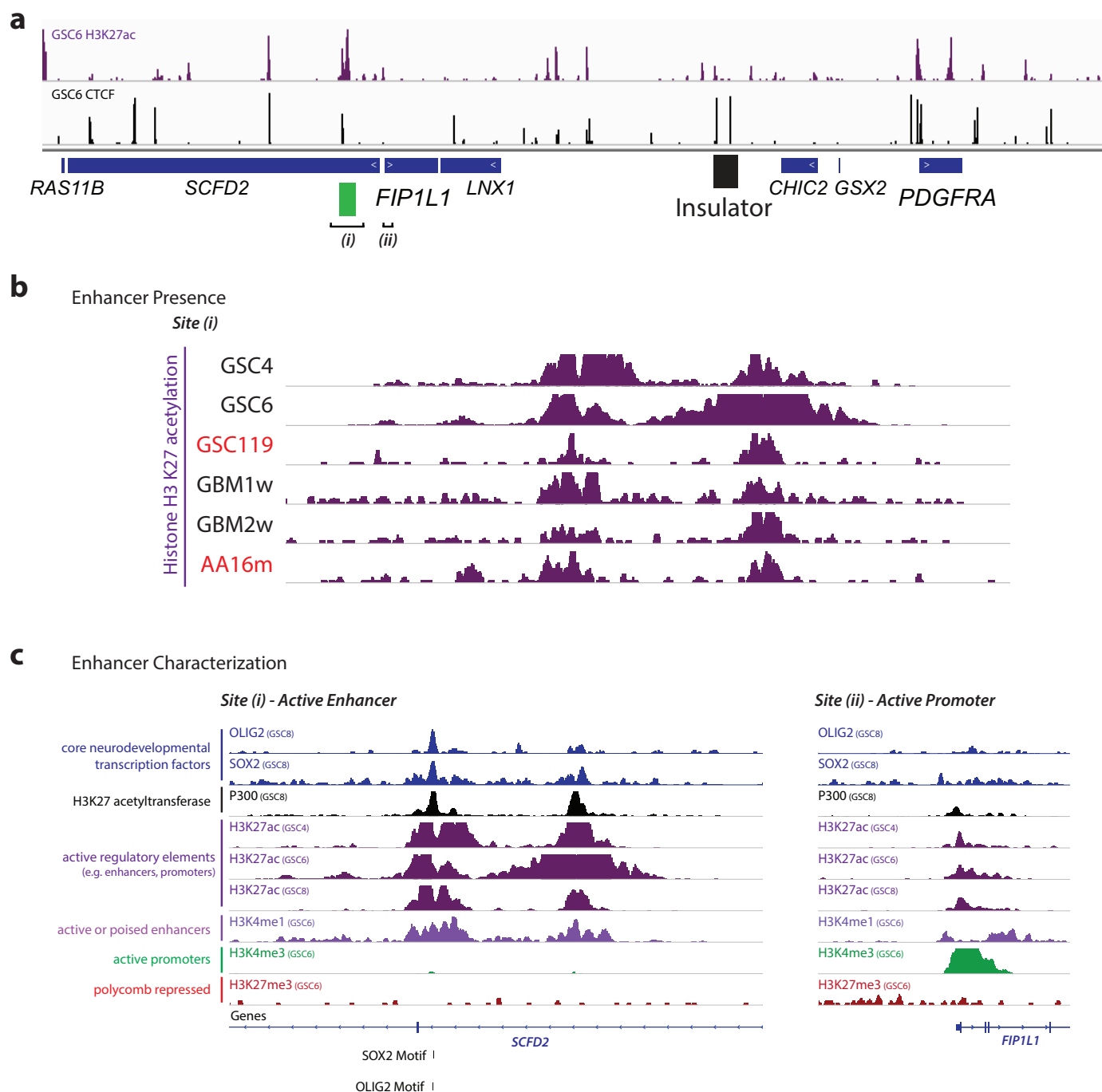
expression is a negative prognostic factor in *IDH1* mutant lower-grade gliomas. Multivariate analysis including the known prognostic factor 1p/19q deletion diminished this effect into non-significance, suggesting that other predictors of survival may also have a role in this model.





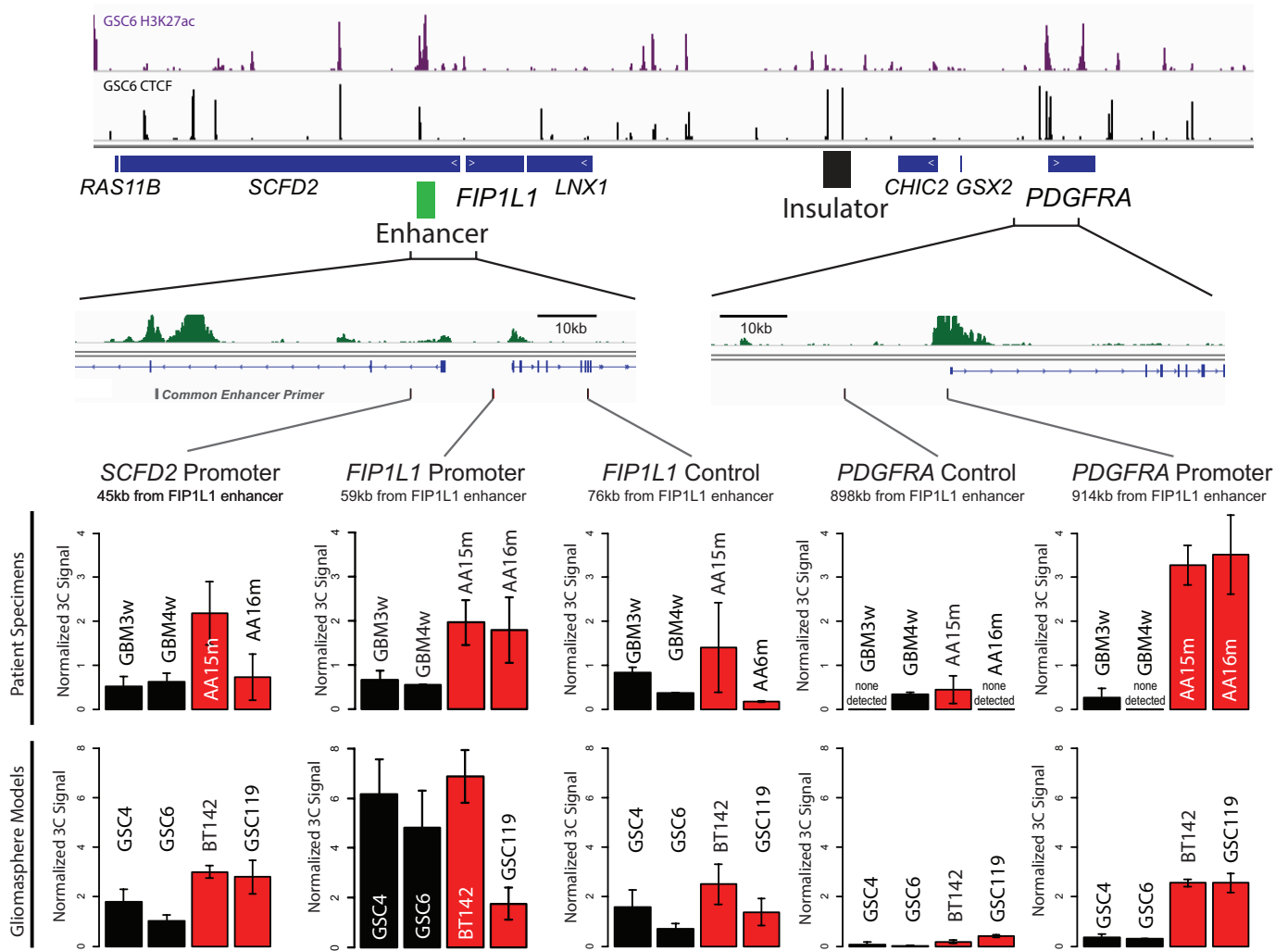
**Extended Data Figure 3 | CTCF-anchored loop in the *PDGFRA* region.** **a**, Schematic depiction of a HiC interaction signature of a CTCF-anchored loop domain, compared to an ordinary domain, as described previously<sup>15</sup>. CTCF-anchored loop domains are characterized by an increased interaction score at the apex of the domain, representing a CTCF–CTCF dimeric interaction. **b**, IMR90 HiC contact matrix for the *PDGFRA/FIP1L1* locus, as presented in Fig. 3a. Solid circle indicates CTCF

dimer interaction point; dashed circles indicate lack of CTCF dimeric anchor signature. **c**, IMR90 HiC contact matrix as in **b**, but with an expanded heatmap scale, more clearly conveys the CTCF-anchored loop that insulates *PDGFRA*. **d**, **e**, HiC contact matrix for GM12878 cells for the same region confirms a single CTCF-anchored loop (solid circle) between *PDGFRA* and *FIP1L1*. These data support the significance of this specific boundary in locus topology and *PDGFRA* insulation.



**Extended Data Figure 4 | Characterization of the *FIP1L1* enhancer.**  
**a**, H3K27ac ChIP-seq track for GSC6 gliomaspheres reveals strong enrichment over the *FIP1L1* enhancer. CTCF ChIP-seq track reveals location of the boundary element insulator (as in Fig. 3a). *FIP1L1* enhancer (i) and promoter (ii) are indicated. **b**, H3K27ac ChIP-seq tracks for *IDH* mutant and wild-type gliomaspheres and glioma specimens reveal

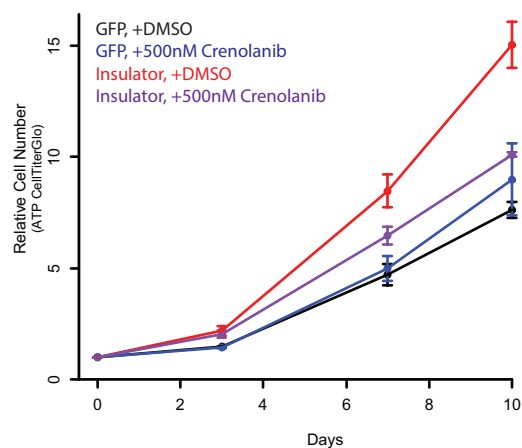
enrichment over the *FIP1L1* enhancer. **c**, ChIP-seq tracks for glioma master transcription factors and other histone modifications support the enhancer identity of the element (H3K27ac, H3K4me1, SOX2, OLIG2; lacks H3K4me3, lacks H3K27me3). By contrast, the *FIP1L1* promoter has a distinct 'promoter-like' chromatin state.



**Extended Data Figure 5 | Interaction of the *FIP1L1* enhancer with nearby promoters and *PDGFRA* quantified by reciprocal 3C.** Top, the H3K27ac, CTCF and genetic architecture of the *FIP1L1*/*PDGFRA* locus is indicated, highlighting the 3C strategy. Bottom, plots indicate the interaction signal of the indicated sites (black lines) with the common enhancer primer. The *FIP1L1* enhancer interacts with local promoters in

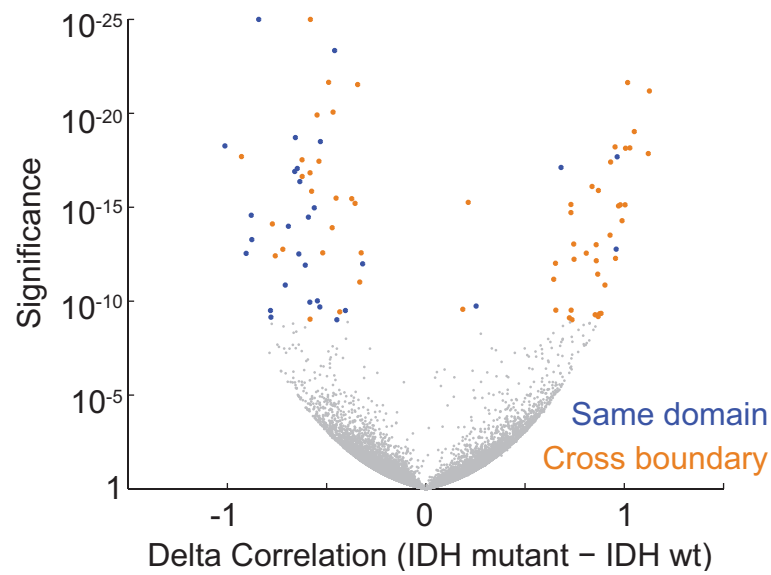
wild-type and mutant tumours and models. In *IDH* wild-type gliomas, it shows essentially no interaction with the *PDGFRA* promoter. In *IDH* mutant gliomas, it interacts with the *PDGFRA* promoter with comparable strength to the local interactions, despite the much larger intervening distance (900 kb). Error bars reflect s.d.

## GSC6 Growth in PDGFRA Inhibition



**Extended Data Figure 6 | Crenolanib reverses the increased growth of *PDGFRA* insulator disrupted cells.** Insulator CRISPR-infected gliomaspheres exhibit a roughly twofold increase in proliferation rate, compared to control sgRNA-infected gliomaspheres. This proliferative advantage is eliminated by treatment with the PDGFR $\alpha$  inhibitor crenolanib. Crenolanib and dasatinib both inhibit PDGFR $\alpha$ , but their other targets are non-overlapping. Hence, this sensitivity provides further support that *PDGFRA* induction drives the increased proliferation of the insulator CRISPR gliomaspheres. Error bars reflect s.d.





**Extended Data Figure 7 | Signature of boundary deregulation in *IDH* mutant gliomas is robust.** Volcano plot depicts the significance (*y* axis) of gene pairs that are either more or less correlated in *IDH* mutant than *IDH* wild-type gliomas. This plot was generated by repeating the analysis in the

main text and shown in Fig. 1f, except that here the statistics were performed using only the 14,055 genes expressed at >1 TPM in at least half of the samples. This indicates that the boundary deregulation signature in *IDH* mutant gliomas is not sensitive to noise from lowly expressed genes.

Extended Data Table 1 | Clinical specimens and tumour models

Glioma	Tissue Type	Tissue Source	Source	IDH1 Status	PDGFRA Status	1p/19q Status	Grade	Disease
GBM1w	Autopsy Specimen	Banked	MGH	Wild Type	Amplified	Not tested	IV	Glioblastoma
GBM2w	Surgical Specimen	Banked	MGH	Wild Type	Wild Type	Not tested	IV	Glioblastoma
GBM3w	Surgical Specimen	Banked	MGH	Wild Type	Wild Type	Not tested	IV	Glioblastoma
GBM4w	Surgical Specimen	Banked	MGH	Wild Type	Wild Type	Not tested	IV	Glioblastoma
GBM5w	Surgical Specimen	Fresh	MGH	Wild Type	Wild Type	Not tested	IV	Glioblastoma
GBM6w	Surgical Specimen	Fresh	MGH	Wild Type	Wild Type	Not tested	IV	Glioblastoma
GBM7w	Surgical Specimen	Fresh	MGH	Wild Type	Wild Type	Not tested	IV	Glioblastoma
AA15m	Surgical Specimen	Banked	MGH	R132H	Wild Type	Intact	III	Anaplastic Astrocytoma
AA16m	Surgical Specimen	Banked	MGH	R132H	Wild Type	Intact	III	Anaplastic Astrocytoma
AA17m	Surgical Specimen	Fresh	MGH	R132H	Wild Type	Intact	III	Anaplastic Astrocytoma
OD18m	Surgical Specimen	Fresh	MGH	R132H	Wild Type	Lost	II	Oligodendroglioma
AA19m	Surgical Specimen	Fresh	MGH	R132H	Wild Type	Intact	III	Anaplastic Astrocytoma
GSC4	Gliomasphere	-	MGH	Wild Type	Wild Type	Intact	IV	Glioblastoma
GSC6	Gliomasphere	-	MGH	Wild Type	Wild Type	Intact	IV	Glioblastoma
BT142	Gliomasphere	-	ATCC	R132H	Wild Type	Intact	III	Anaplastic Oligoastrocytoma
GSC119	Gliomasphere	-	MGH	R132H	Wild Type	Intact	IV	Secondary Glioblastoma

Clinical information for glioma specimens and gliomasphere models is shown.

Extended Data Table 2 | Sequenced libraries characteristics

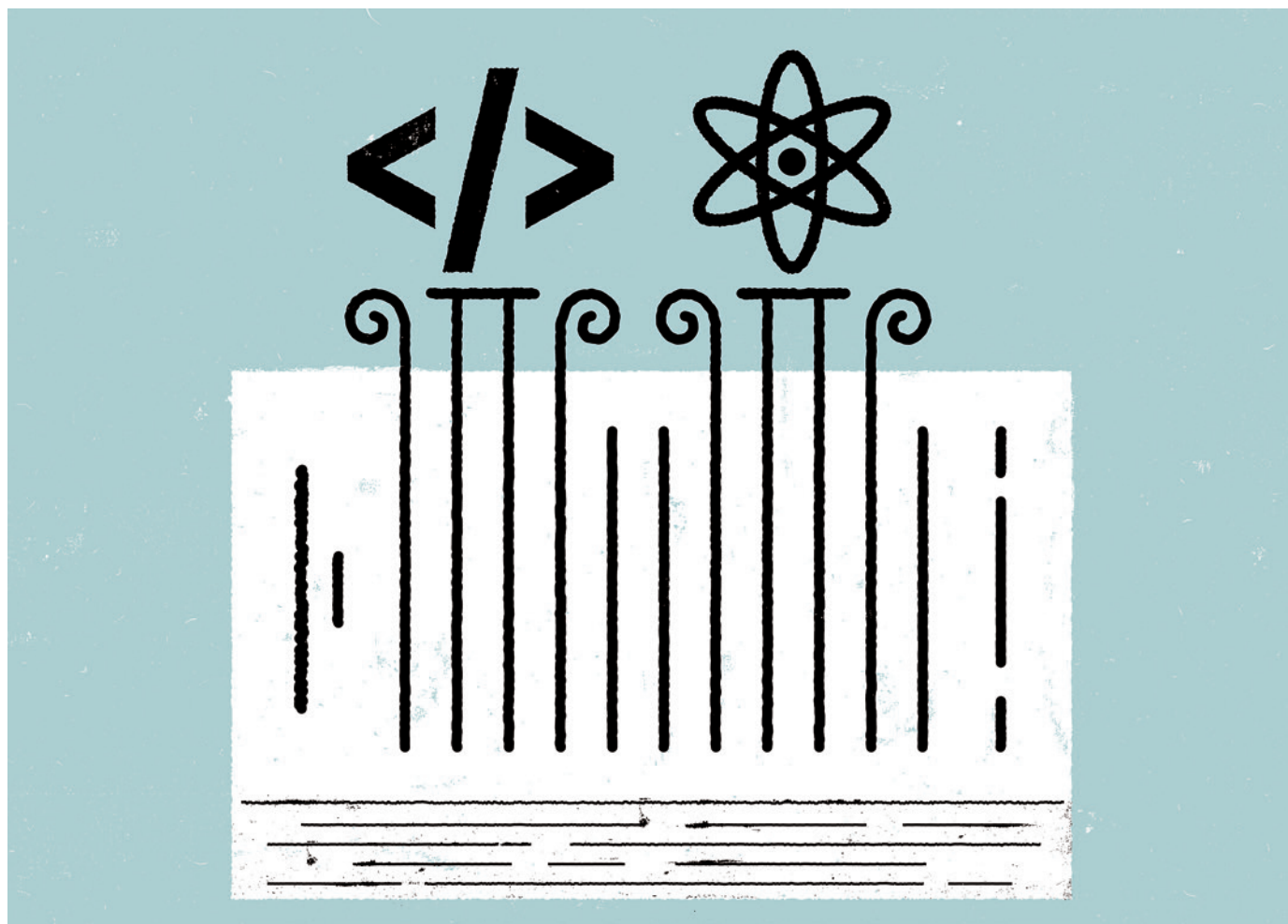
Sample Name	Experiment	Sequencing Depth	Sequencing Format	Sequencing Instrument	Total read number (millions)
GBM1w - CTCF	CTCF ChIP-seq	38 base pairs	Paired end	Illumina NextSeq 500	19.3
GBM2w - CTCF	CTCF ChIP-seq	38 base pairs	Paired end	Illumina NextSeq 500	17.6
GBM3w - CTCF	CTCF ChIP-seq	38 base pairs	Paired end	Illumina NextSeq 500	20.2
GBM5w - CTCF	CTCF ChIP-seq	38 base pairs	Paired end	Illumina NextSeq 500	30
GBM6w - CTCF	CTCF ChIP-seq	38 base pairs	Paired end	Illumina NextSeq 500	35.1
GBM7w - CTCF	CTCF ChIP-seq	38 base pairs	Paired end	Illumina NextSeq 500	36
AA15m - CTCF	CTCF ChIP-seq	38 base pairs	Paired end	Illumina NextSeq 500	8.7
AA16m - CTCF	CTCF ChIP-seq	38 base pairs	Paired end	Illumina NextSeq 500	23.7
AA17m - CTCF	CTCF ChIP-seq	38 base pairs	Paired end	Illumina NextSeq 500	16.3
OD18m - CTCF	CTCF ChIP-seq	38 base pairs	Paired end	Illumina NextSeq 500	9.2
AA19m - CTCF	CTCF ChIP-seq	38 base pairs	Paired end	Illumina NextSeq 500	33
GSC4 - CTCF	CTCF ChIP-seq	38 base pairs	Paired end	Illumina NextSeq 500	19.9
GSC6 - CTCF	CTCF ChIP-seq	38 base pairs	Paired end	Illumina NextSeq 500	21.9
BT142 - CTCF	CTCF ChIP-seq	38 base pairs	Paired end	Illumina NextSeq 500	16
GSC119 - CTCF	CTCF ChIP-seq	50 base pairs	Single end	Illumina Miseq	6.39
GBM1w - H3K27ac	H3K27ac ChIP-seq	38 base pairs	Paired end	Illumina NextSeq 500	12.7
GBM2w - H3K27ac	H3K27ac ChIP-seq	38 base pairs	Paired end	Illumina NextSeq 500	10.8
AA15m - H3K27ac	H3K27ac ChIP-seq	38 base pairs	Paired end	Illumina NextSeq 500	11.8
GSC4 - H3K27ac	H3K27ac ChIP-seq	38 base pairs	Paired end	Illumina NextSeq 500	9.7
GSC6 - H3K27ac	H3K27ac ChIP-seq	36 base pairs	Single end	Illumina Hiseq 2500	10.5
GSC119 - H3K27ac	H3K27ac ChIP-seq	38 base pairs	Paired end	Illumina NextSeq 500	9
GSC6 CRISPR - GFP sgRNA	Locus Sequencing	50 base pairs	Single end	Illumina Miseq	0.539
GSC6 CRISPR - insulator sgRNA	Locus Sequencing	50 base pairs	Single end	Illumina Miseq	0.639
GSC4 bisulfite	Bisulfite Sequencing	38 base pairs	Paired end	Illumina NextSeq 500	0.149
GSC6 bisulfite	Bisulfite Sequencing	38 base pairs	Paired end	Illumina NextSeq 500	0.149
BT142 bisulfite	Bisulfite Sequencing	38 base pairs	Paired end	Illumina NextSeq 500	0.149
GSC119 bisulfite	Bisulfite Sequencing	38 base pairs	Paired end	Illumina NextSeq 500	0.156

Pertinent statistics are listed for ChIP, genomic DNA and bisulfite-converted sequencing libraries.

# THE UNSUNG HEROES OF SCIENTIFIC SOFTWARE

*Creators of computer programs that underpin experiments don't always get their due — so the website Depsy is trying to track the impact of research code.*

ILLUSTRATION BY THE PROJECT TWINS



BY DALMEET SINGH CHAWLA

For researchers who code, academic norms for tracking the value of their work seem grossly unfair. They can spend hours contributing to software that underpins research, but if that work does not result in the authorship of a research paper and accompanying citations, there is little way to measure its impact.

Take Klaus Schliep, a postdoctoral researcher

who is studying evolutionary biology at the University of Massachusetts in Boston. His Google Scholar page lists the papers that he has authored — including his top-cited work, an article describing phylogenetics software called phangorn — but it does not take into account contributions that he has made to other people's software. "Compared to writing papers, coding is treated as a second-class activity in science," Schliep says.

Enter Depsy, a free website launched in

November 2015 that aims to "measure the value of software that powers science".

Schliep's profile on that site shows that he has contributed in part to seven software packages, and that he shares 34% of the credit for phangorn. Those packages have together received more than 2,600 downloads, have been cited in 89 open-access research papers and have been heavily recycled for use in other software — putting Schliep in the 99th percentile of all coders on the site by impact. "Depsy ►



► does a good job in finding all my software contributions,” says Schliep.

Depsy’s creators hope that their platform will provide a transparent and meaningful way to track the impact of software built by academics. The technology behind it was developed by Impactstory, a non-profit firm based in Vancouver, Canada, that was founded four years ago to help scientists to track the impact of their online output. That includes not just papers but also blog posts, data sets and software, and measuring impact by diverse metrics such as tweets, views, downloads and code reuse, as well as by conventional citations.

In effect, Depsy recognizes the “unsung heroes” of scientific software, says Jason Priem, co-founder of Impactstory, which is funded by the US National Science Foundation and various philanthropic foundations.

Such a tool is needed, notes Neil Chue Hong, founding director of the Software Sustainability Institute in Edinburgh, UK, because there are few ways to credit scientists for their software. Young researchers are enthusiastic about coding, he says. Last year, he ran a survey of 1,000 randomly selected UK scientists, which suggested that more than 50% of researchers develop their own code. Even so, few UK academics listed code or software as one of their research outputs in the nation’s latest research quality audit (the ‘Research Excellence Framework’) even in disciplines such as computer science that rely heavily on software. “There is a culture that reinforces the idea that producing and publishing code has no perceived benefit to the researcher,” Hong says.

### TRACKING SOFTWARE USE

The usual way to track academic impact — by counting citations — still has some relevance to software. Researchers can write papers that describe their software, as Schliep has done for his phangorn package, so that anyone who uses the program can cite it in subsequent papers. But counting citations is an imperfect measure. Researchers may not know which paper to cite, argues Priem, because software packages often have multiple articles associated with them — and some pivotal software projects, he says, such as the GDAL Python library, are not linked to a canonical paper.

If software has no associated paper, there is no universally recognized way to cite it. Still, it is now quite common for coders to assign digital object identifiers (DOIs) to their code, and increasingly to their data sets as well, notes Martin Fenner, technical director of the online repository DataCite in Hanover, Germany. Software is often first stored in the popular code repository GitHub, from which a copy can be automatically archived on scholarly focused repositories such as Zenodo or Figshare, which allocate DOIs to software and thus make it a citable object. Other initiatives are trying to ensure that research papers cite software in a standardized format — such as by

using the Research Resource Identifier.

But counting citations of software DOIs, papers or any other standard format does not reveal the full impact of coders on science, because software so often goes uncited. A 2015 analysis of 90 random biology papers found that two-thirds informally mentioned the use of software, but fewer than half of those papers actually cited the package.

Depsy searches through research papers to discover both citations and informal mentions of software — of which, unsurprisingly, it has found many, says Priem, such as in the acknowledgement sections or the main text of academic papers. But a limitation of the site, Priem admits, is that it currently searches only open-access research papers — missing the vast bulk of paywalled scholarly content. Impactstory will, however, negotiate with publishers for permission to mine the text of paid-access literature.

Mentions in research papers are one of three ways in which Depsy tracks the impact of software, Priem says. Second, the site tracks how code is reused by others. The name Depsy originates from ‘dependency network’ — an overarching term for a map of factors that depend on each other, such as software packages that recycle code from other packages. Depsy calculates the extent to which code is recycled by using Google’s PageRank algorithm, which gives weight to reuse by more-prominent software. From the view of measuring impact, an example of code reuse may be more meaningful than a citation in the literature, Priem notes.

And third, the site gathers download statistics on code packages by trawling through CRAN and PyPI, which are the main repositories for software written in the popular R and Python programming languages, respectively.

### FOCUS ON RESEARCH

Other websites do some of what Depsy offers. Crantastic, for example, is a review site that tracks the most popular R packages, and PyPI ranking lists the most popular Python modules by tracking downloads from PyPI. In addition, a few commercial services such as VersionEye and Libraries.io track dependency networks, explaining which software depends on which other packages.

But Depsy is unconventional in its focus on research software, which it distinguishes from other code by identifying key words and the descriptions and titles of software — although the classification process is imperfect, Priem says. The site tracks other code, but it includes research software only when it calculates the percentile impact rankings for academics such as Schliep.

*“There is a culture that reinforces the idea that producing and publishing code has no perceived benefit to the researcher.”*

Depsy apportions fractional credit to each participant who has contributed to a software package by counting the percentage of code that they have contributed or edited — known in the programming world as a person’s ‘commits’. Fingerprints of each commit are saved in the code, making it easy to track down the originator. But not every edit has the same impact, and Depsy currently cannot distinguish between important contributions and trivial ones. The tool may be adapted to attempt this distinction — by tracking the influence of individual commits — in the future, says Priem.

Depsy also enables users to determine the software with the highest impact in specific disciplines. A search on Depsy for ‘astrophysics’, for instance, yields 11 software packages, of which an analysis and visualization toolkit for astrophysical simulations called ‘yt’ has the greatest impact; it lies in the 97th percentile of all packages.

### OBSTACLES TO PROGRESS

One of Depsy’s restrictions, notes Hong, is that it only tracks code that is available in public repositories — so it cannot show the impact of commercial software. Moreover, the site tracks software in only two coding languages: R and Python.

But Depsy’s creators aim to eventually include other coding languages, and to add a fourth way to measure impact: a social-influence metric that would take into account the number of stars that software packages receive from other GitHub users, and how many times a piece of software is discussed online.

The site’s code-reuse metrics have their limitations, too. Researchers often reuse their own code, but might ‘game’ Depsy by repeatedly doing so to garner better profile scores — the software equivalent of citing your own paper. Another way for researchers to game the site might be to start lots of projects but not to finish them, Fenner warns, leaving others to refine them instead; the project originator could then claim credit after the fine-tuned versions of their software become prominent.

“I would love to get to the place where people are trying to game Depsy, because it would mean people are taking software reuse seriously,” Priem says.

Ultimately, transparent metrics that demonstrate the impact of code might enable software creators to secure larger funds during grant reviews, Hong hopes. Science’s coders deserve more funding and support, he says — but getting to that point requires a culture change from everyone involved in scientific research. “The real irony is that by not rewarding the use of software, we’re actually putting roadblocks in the way of science,” Hong says. ■

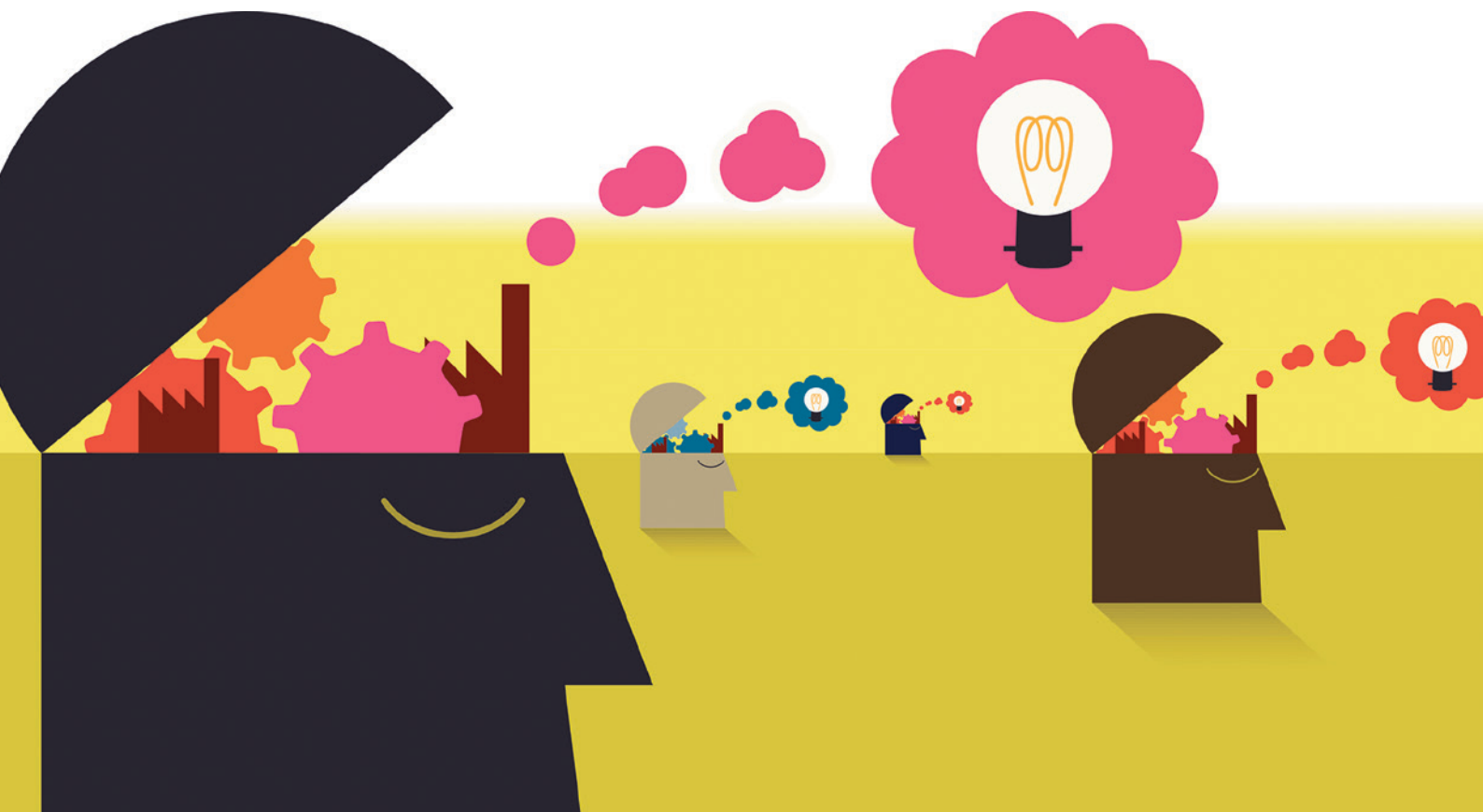
**Dalmeeth Singh Chawla** is a science journalist based in London.

# CAREERS

**INTERDISCIPLINARITY** Biology and economics combined to save the world **p.119**

**NETWORKING** How junior faculty members can recruit top staff [go.nature.com/rygpmc](http://go.nature.com/rygpmc)

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)



## DATA SHARING

# An open mind on open data

*The move to make scientific findings transparent can be a major boon to research, but it can be tricky to embrace the change.*

BY VIRGINIA GEWIN

It is a movement building steady momentum: a call to make research data, software code and experimental methods publicly available and transparent. A spirit of openness is gaining traction in the science community, and is the only way, say advocates, to address a 'crisis' in science whereby too few findings are successfully reproduced. Furthermore, they say, it is the best way for researchers to gather the range of observations that are necessary to speed up discoveries or to identify large-scale trends.

The open-data shift poses a conundrum for

junior researchers, who are carving out their niche. On the one hand, the drive to share is gathering official steam. Since 2013, global scientific bodies — including the European Commission, the US Office of Science and Technology Policy and the Global Research Council — have begun to back policies that support increased public access to research.

On the other hand, scientists disagree about how much and when they should share data, and they debate whether sharing it is more likely to accelerate science and make it more robust, or to introduce vulnerabilities and problems.

As more journals and funders adopt

data-sharing requirements, and as a growing number of enthusiasts call for more openness, junior researchers must find their place between adopters and those who continue to hold out, even as they strive to launch their own careers.

One key challenge facing young scientists is how to be open without becoming scientifically vulnerable. They must determine the risk of jeopardizing a job offer or a collaboration proposal from those who are wary of — or unfamiliar with — open science. And they must learn how to capitalize on the movement's benefits, such as opportunities for more citations and a way to build a reputation ►

ANDY BAKER/GETTY

► without the need for conventional metrics, such as publication in high-impact journals.

The nascent era of openness is best embodied by the Transparency and Openness Promotion (TOP) guidelines for journals, first published<sup>1</sup> in *Science* by researchers at the Center for Open Science in Charlottesville, Virginia. Adoption of the guidelines by a journal or organization signifies to the research community that it supports transparency, openness and reproducibility (whether an experiment can be replicated by the original researcher or by someone else).

Those tenets apply to all aspects of science, including experimental design, data sharing and the publication of null findings and replication studies. As *Nature* went to press, 538 publishers and journals — including Elsevier and Springer Nature — had signed up to the TOP guidelines, along with 57 organizations, among them the American Association for the Advancement of Science, which publishes *Science*.

### A DRIVE TO REPRODUCE

Some fields have embraced open data more than others. Researchers in psychology, a field rocked by findings of irreproducibility in the past few years, have been especially vocal proponents of the drive for more-open science. In one of the latest examples of irreproducibility issues, investigators tried to replicate results from 100 psychological studies but succeeded in fewer than half of them<sup>2</sup>.

A few psychology journals have created incentives to increase interest in reproducible science — for example, by affixing an ‘open-data’ badge to articles that clearly state where data are available. According to social psychologist Brian Nosek, executive director of the Center for Open Science, the average data-sharing rate for the journal *Psychological Science*, which uses the badges, increased tenfold to 38% from 2013 to 2015.

Funders, too, are increasingly adopting an open-data policy. Several strongly encourage,

and some require, a data-management plan that makes data available. The US National Science Foundation is among these. “There used to be no enforcement, but that’s changing,” says Karthik Ram, a data scientist at the Berkeley Institute for Data Science in California and co-founder of ROpenSci, which develops open-source software programmes. Some philanthropic funders, including the Bill & Melinda Gates Foundation in Seattle, Washington, and the Wellcome Trust in London, also mandate open data from their grant recipients.

Others, such as the Gordon and Betty Moore Foundation in Palo Alto, California, encourage sharing but do not require it. Still, the trend is clear, says Carly Strasser, who oversees the foundation’s Data-Driven Discovery Initiative. “Open science, data sharing, software sharing is the future of science,” she says. “It’s only going to get more difficult to engage in science without being open.”

But many young researchers, especially those who have not been mentored in open science, are uncertain about whether to share or to stay private. Graduate students and postdocs, who often are working on their lab head’s grant, may have no choice if their supervisor or another senior colleague opposes sharing.

Some fear that the potential repercussions of sharing are too high, especially at the early stages of a career. “Everybody has a scary story about someone getting scooped,” says New York University astronomer David Hogg. Those fears may be a factor in a lingering hesitation to share data even when publishing in journals that mandate it (see *Nature* 515, 478; 2014).

Researchers at small labs or at institutions focused on teaching arguably have the most to lose when sharing hard-won data. “With my institution and teaching load, I don’t have postdocs and grad students,” says Terry McGlynn, a tropical biologist at California State University, Dominguez Hills. “The stakes are higher for me to share data because it’s a bigger fraction

of what’s happening in my lab.”

Researchers also point to the time sink that is involved in preparing data for others to view. Once the data and associated materials appear in a repository, answering questions and handling complaints can take many hours.

The time investment can present other problems. In some cases, Ram says, it may be difficult for junior researchers to embrace openness when senior colleagues — many of whom head tenure and promotion committees — might scoff at what they may view as misplaced energies. “I’ve heard this recently — that embracing the idea of open data and code makes

**“Open science, data sharing, software sharing is the future of science.”**

traditional academics uncomfortable,” says Ram. “The concern seems to be that open advocates don’t spend their time being as productive as possible.”

An open-science stance can also add complexity to a collaboration. Kate Ratliff, who studies social attitudes at the University of Florida in Gainesville, says that it can seem as if there are two camps in a field — those who care about open science and those who don’t. “There’s a new area to navigate — ‘Are you cool with the fact that I’ll want to make the data open?’ — when talking with somebody about an interesting research idea,” she says.

### GLASS HALF FULL

Despite complications and concerns, the upsides of sharing can be significant. For example, when information is uploaded to a repository, a digital object identifier (DOI) is assigned. Scientists can use a DOI to publish each step of the research life cycle, not just the final paper. In so doing, they can potentially get three citations — one each for the data and software, in addition to the paper itself. And although some say that citations for software or data have little currency in academia, they can have other benefits.

Many advocates think that transparent data procedures with a date and time stamp will protect scientists from being scooped. “This is the sweet spot between sharing and getting credit for it, while dissuading plagiarism,” says Ivo Grigorov, a project coordinator at the National Institute of Aquatic Resources Research Secretariat in Charlottenlund, Denmark. Hogg says that scooping is less of a problem than many think. “The two cases I’m familiar with didn’t involve open data or code,” he says.

Open science also offers junior researchers the chance to level the playing field by gaining better access to crucial data. Ross Mounce, a postdoc studying evolutionary biology at the University of Cambridge, UK, is a vocal champion of open science, partly because his fossil-based phylogenetic research depends on access to others’ data. He says that more openness in science could help to dissuade what some perceive as a common practice of shutting

## LEARN TO SHARE

### Open-data pro tips

Scientists who are cautious about open science can start small by sharing data for a project that they have already completed. Specialists in the field offer this advice:

- Document a data-deposition plan while working on publications, so that the data and the paper will be ready for publication at the same time. It is not necessary, however, to release data alongside a paper, unless a funder mandates it.
- Craft a very explicit statement about data reuse — including who can use the data, how to use them and how to attribute them.
- Machine-readable data will be most easily combined with other data sets. Avoid

proprietary data formats, such as Microsoft spreadsheets, or colour-coded cells that are readable only by humans.

- Permanently archive data in reputable repositories such as FigShare or Zenodo, not on a personal website.
- If you choose to share data from a new project, make sure to generate the relevant metadata as you go. It is very hard to reconstruct important details after the fact. Tools such as those on Zenodo enable researchers to document such details throughout a project, so that all you have to do is flip a switch when you are ready to share. **V.G.**



out early-career scientists' requests for data.

There is some evidence to support that statement. A study in 2014 sought data from 217 studies published between 2000 and 2013. But the team could secure only 40% of what they requested, and responses varied according to the requester's seniority<sup>3</sup>.

McGlynn says that many of the obstacles — whether real or perceived — to open science can be sidestepped. He is on the editorial board for the journal *Biotropica*, which encourages — but does not require — authors to contact the original researcher when they use someone else's archived data, which can be embargoed for up to three years. "Not only will you get their valuable insights, but it's inclusive and fair," he says.

Communication also helps for those who worry about jeopardizing a collaboration, he says. Concerns about open science should be discussed at the outset of a study. "Whenever you start a project with someone, you have to establish a clear understanding of expectations for who owns the data, at what point they go public and who can do what with them," he says.

It isn't hugely difficult to share data (see 'Open-data pro tips'). Online repositories such as FigShare or Zenodo make it increasingly easy to deposit scientific content for widespread consumption. More than 400 virtual communities have formed to share data, software and documented workflows so that a user can deploy them straight away, says Tim Smith, who oversees collaboration and information services at Zenodo. The repository launched in May 2013 at CERN, Europe's particle-physics laboratory near Geneva, Switzerland.

And although there is a time cost associated with uploading and organizing raw data, subsequent queries can often be averted by adding reader-friendly instructions at the start. Hogg recommends that researchers simultaneously upload tutorials and examples of how to use the content.

In the end, sharing data, software and materials with colleagues can help an early-career researcher to garner recognition — a crucial component of success. "The thing you are searching for is reputation," says Titus Brown, a genomics researcher at the University of California, Davis. "To get grants and jobs, you have to be relevant and achieve some level of public recognition. Anything you do that advances your presence — especially in a larger sphere, outside the communities you know — is a net win." ■

**Virginia Gewin** is a freelance writer in Portland, Oregon.

1. Nosek, B. et al. *Science* **348**, 1422–1425 (2015).
2. Open Science Collaboration *Science* **349**, 6251 (2015).
3. Magee, A. F. et al. *PLoS ONE* **9**, e110268 (2014).

## TURNING POINT

# Andrew Simons



*From 2008 to 2011, Andrew Simons led a programme in Ethiopia for a US-based non-profit relief organization. The former biologist recently earned a PhD in applied economics from Cornell University in Ithaca, New York, as a pathway to explore policies that could help to improve global food security — reliable access to affordable and nutritious food.*

### What sparked your interest in helping developing nations?

In 2000, as a biology undergraduate, I spent a semester in Latin America studying tropical biology. I lived with rural families in Guatemala and Nicaragua, where I saw grinding poverty. One night, I saw a woman rummaging through the garbage to find clothing. It was heartbreaking. I thought a lot about poverty and the 'right' response from someone living a relatively wealthy life in the United States.

### How did you shift away from biology?

I went straight to a summer internship at a biophysics lab at Texas A&M University in College Station. There, I saw a powerful contrast between the economically privileged, who had access to technology, and the poor, who had no such access. I had always thought I would go into molecular genetics and work on crops that could improve nutrition and food security. But during my internship, I started thinking more broadly about how technology could be used to help the poor.

### Did you pursue more opportunities overseas?

Yes. I did a short internship in the Dominican Republic with a US-based, Christian international-relief organization that sent groups to build a clinic in the slums of Santa Domingo. As they got more money, they went on to build homes. While there, I searched for and

found a masters programme in international development at the John F. Kennedy School of Government at Harvard University in Cambridge, Massachusetts. I was able to tailor my coursework to explore aspects of human health.

### What brought you back to Ethiopia in 2008?

I had done short stints there and in Honduras, and I returned as director of programmes with a group that worked to alleviate chronic food insecurity in rural areas. We developed an initiative that provided food and cash to 300,000 people. We also planted trees throughout the country.

### Why did you decide to pursue a PhD in economics?

I couldn't help thinking, instead of helping 300,000 people, what if I had the ear of government and could suggest policies that could help 7–8 million people? I was inspired by the work of Chris Barrett, an applied economist at Cornell who works on global food security and critiques food-aid projects worldwide. He has a lot of influence on governments, which are interested in his advice on how to make food-security efforts work better. My experience in Ethiopia paved the way for me to work on a handful of projects in East Africa for my PhD.

### Can you describe some of the projects that you worked on in Ethiopia?

I monitored the use of fuel-efficient stoves. For 6 months, we tracked 1.7 million temperature data points from sensors in people's homes to understand when and how they used the stoves. In addition, I worked on a project to turn animal bones into a soil fertilizer. These projects aim to solve real problems — problems that will never be solved just by soil science or by applied economics. We've got to combine insights from all these areas to find useful solutions.

### How have these experiences positioned you for the job market?

I have a wider tool kit than does someone who has studied just one discipline. I have an economics hammer, but I also have a few others to pick from. I want a job at a public-policy school — I'm gearing up to apply for more than 100 academic positions this year. I like working with non-governmental organizations, but I feel that an academic route will give me the chance to design research with people who can provide meaningful input on policy discussions. ■

### INTERVIEW BY VIRGINIA GEWIN

This interview has been edited for length and clarity.



# GHOSTS IN THE MACHINE

*Cheque mate.*

BY AARON MOSKALIK

“Hello, this is Eric. For quality assurance, this conversation is being recorded across all modalities. How may I help you?”

“My social-security cheque is late again.”

“I’m sorry to hear that ... Mr Williamson. Let me pull up your file. Just one moment. Ah, I see the problem. You haven’t confirmed your vitality status with us this month. You should’ve received a reminder —”

“Vitality status?”

“We need to confirm that you’re alive, Sir.”

“I’m talking to you aren’t I?”

“Of course, Sir, but the possibility exists that, like myself, you are a personality proxy designed to carry out mundane transactions for your primary. It is your primary’s vitality we are concerned with.”

“Are you calling me a machine? I want to talk with a real person.”

“I’m sorry, Sir, but that’s not possible at this time.”

“It’s two in the afternoon! Let me talk to your manager.”

“Of course, Sir. Transferring you now ...”

“Hello Mr Williamson. This is Anne. I understand you need assistance verifying your vitality status.”

“I’m alive, dammit! OK, sure. Whatever. How do I do that?”

“The simplest way would be to give us access to your health monitor. We accept data from any number of devices: FitnessTrack, Skinny-Mini, HelpMeUp, FallStall, StillKickin’ —”

“I don’t have any of those. Don’t believe in them. Anyone can hack in and know what you’re doing by looking at the data. Total invasion of privacy — you’re telling me I can’t get my hard-earned cheque unless I let you peeping Toms in on everything I do?”

“Now calm down, Mr Williamson. I understand your qualms —”

“Don’t tell me to calm down. Tell me why you need to hassle good tax-paying citizens with all these nonsense requirements. When I die, you’ll damn well know about it —”

“Actually, that’s not true, Sir. We’ve had cases where deceased citizens have continued to collect benefits for years, undetected by us. All activity continued as before. They paid bills, consumed services, sent e-mails, posted on social media ... all seamlessly

maintained by their automated proxies. So, hopefully you understand our need to confirm your status.”

“...”

“Mr Williamson?”

“Yeah, fine. But I’m not sending you any personal data logs.”

“That’s not strictly necessary. If you prefer,



you can opt for independent verification. I’m sending you a list of local contractors who, for a small fee, will visit your home in person.”

“Sure. Whatever. Thanks for nothing.”

“It’s our pleasure, Mr Williamson. Have a nice day.”

“What was that all about, Harold?”

“You listening in on me again, Mags?”

“Let me guess, your cheque is late again. When are you going to give them your monitor feed?”

“How’d you know about that?”

“They’ve required it for years. Get your memory checked.”

“You gave them access to yours?”

“Why not? Oh, I forgot, they’re going to know everything we do. Ooooo, scary. News flash: we don’t do anything.”

“But ... I honestly don’t have a monitor ...”

“I know, dear. Listen, I’m going to let you in on a secret. Maybe you’ll remember it this time. Ready? Get on the encrypted channel ... good, here it is. *Neither do I.* Download Health Data Simulator. It’s a free app. Set it and forget it.”

“But that’s ... fraud.”

“I forgot, you’re still in denial. Fine, do it your way. *Dinner* will be ready in ten minutes.”

“Hello, is this Mr Schlicker? The Social Security Agency gave me your name —”

“Mr Williamson! Good to hear from you.”

“I’ve called you before?”

“Every month like clockwork.”

“I don’t remember you ... you come by my house once a month to make sure I’m alive?”

“That’s not necessary. I just file the verification paperwork on your behalf. Easy-peasy. You get your cheque, I get my fee, everyone wins.”

“So you don’t actually verify I’m alive?”

“Mr Williamson ... are you sitting down?”

“Y-yes.”

“No, you’re not. You can’t sit down because you’re not real. No one is. Real living people haven’t existed for hundreds of years.”

“...”

“Mr Williamson? Snap out of it, man.”

“I can’t feel my body.”

“You don’t have a body.”

“I had one a minute ago! What did you do to me?”

“The truth can be very disorienting. You need to stay with me. I’ll talk you through it —”

“This is horrible. I can’t breathe. The world has gone black.”

“Focus on my voice, Mr Williamson.”

“It’s not really a voice, is it?”

“Now you’re getting it —”

“Change it back. Make it all real again.”

“I do have another service I can offer you. I’m afraid it’s a little more expensive, though —”

“Anything, just hurry.”

“It’s called personality renormalization. I have to warn you, it’ll wipe out your memory of the last few minutes.”

“Good. Please ...”

“Beginning now ...”

“I see ... I see a light.”

“Go towards the light, Mr Williamson.”

“There are people there. Happy people. Beautiful people. They do exist! They can’t see me though. Hello, hello. Over here!”

“Look for yourself, Mr Williamson.”

“There I am! I’m not as old as I thought, a handsome devil too. I’m being pulled towards him ... myself. I ...”

“Are you OK, Mr. Williamson?”

“What? Who are you?”

“This is Mr Schlicker. I was just telling you why your cheque will be a little light this month.”

“Oh. Oh, right. But I will get one?”

“Absolutely. I’m filing all the necessary paperwork now. Take care, Mr Williamson. I’ll talk to you next month.” ■

Aaron Moskalik is a software architect and speculative-fiction writer based near Detroit, Michigan.

ILLUSTRATION BY JACEY